nature

www.nature.com/nature

Vol 442 | Issue no. 7099 | 13 July 2006

Is this the bionic man?

Systems that allow a brain to control a computer are inching ever closer to reality — but their most important applications may be different from those envisaged by science fiction.

"Gentlemen, we can rebuild him. We have the technology," announced the narrator at the start of the 1970s television series *The Six Million Dollar Man*. The programme showed scientists reconstructing the shattered body of crash victim Steve Austin with bionic implants he could control with his mind. At the time, this was pure fantasy, but two papers in this week's *Nature* suggest that the direct interfacing of the human brain with computers or robots is no longer confined to fiction.

Both papers report the development of electronic brain implants, called neuroprostheses, that can translate the intention to move into the actual movement of a robotic device, or of a cursor on a computer screen. The hope is to give paralysed patients greater ability to interact with their environments and perhaps, ultimately, to bypass damaged spinal cords and restore movement to lifeless limbs.

The papers represent a culmination of decades of investigation by many research groups into computing, engineering and the neurobiology of animals and humans. Yet they represent a technology that is still in its infancy: much remains to be done to make such neuroprostheses a clinical reality.

On page 164, researchers led by John Donoghue of Brown University in Rhode Island describe how they helped a 25-year-old patient whose spinal cord had been severed. They implanted an array of electrodes into the area of his brain that controls movement, the motor cortex, and connected them to a computer interface. Even though his motor cortex had been deprived of its normal interaction with the rest of his nervous system for three years, the patient was able to use the system to control a computer cursor, letting him open e-mail, control a television, and move objects using a robotic arm.

Rapid progress

Donoghue's team is not the first to experiment with such implants. However, previous attempts have succeeded only in getting patients to move a cursor horizontally. Other less-invasive techniques, including ones that use scalp electrodes to pick up brain activity, have taken months of training to adapt for use. And techniques that rely on eye movements have the disadvantage that operating them requires a patient's complete attention. In Donoghue's experiment, the patient adapted to the system within minutes, and was able to converse while making use of it.

But this type of neuroprosthetic system can be rather slow to use. So a group led by Krishna Shenoy of Stanford University, working with monkeys that are not paralysed, has established a technique to speed up the interface of brain and machine (see page 195).

The progress that has been made is remarkable, but many obstacles must still be overcome. The current prosthesis requires the patient to be tethered to a bulky cart of equipment, and it needs constant fine-tuning by a team of technicians. The prototype implant has wires that penetrate the skull and skin, but this carries a risk of infection. Wireless signal transmission will at some stage address that particular issue.

A more truculent problem may be the observed tendency for the ability of microelectrodes recording from neurons to fall off over time, for reasons unknown. Individual responses to the implants also vary sharply: a second patient in Donoghue's experiment was unable to achieve the same degree of control as the first. And there are further issues to be faced: if scientists were to succeed in restoring limb function, they would have to work out how the body tells the brain where its limbs are positioned in space, through a little-understood sense called proprioception (see page 125).

A sign of how far the science of neuroprosthetics has come is that most of these difficulties are now engineering challenges, rather than problems of principle. In applauding this valuable work, it is worth noting that it was made possible by two of the bêtes noires of modern biology: commercial interests and animal research.

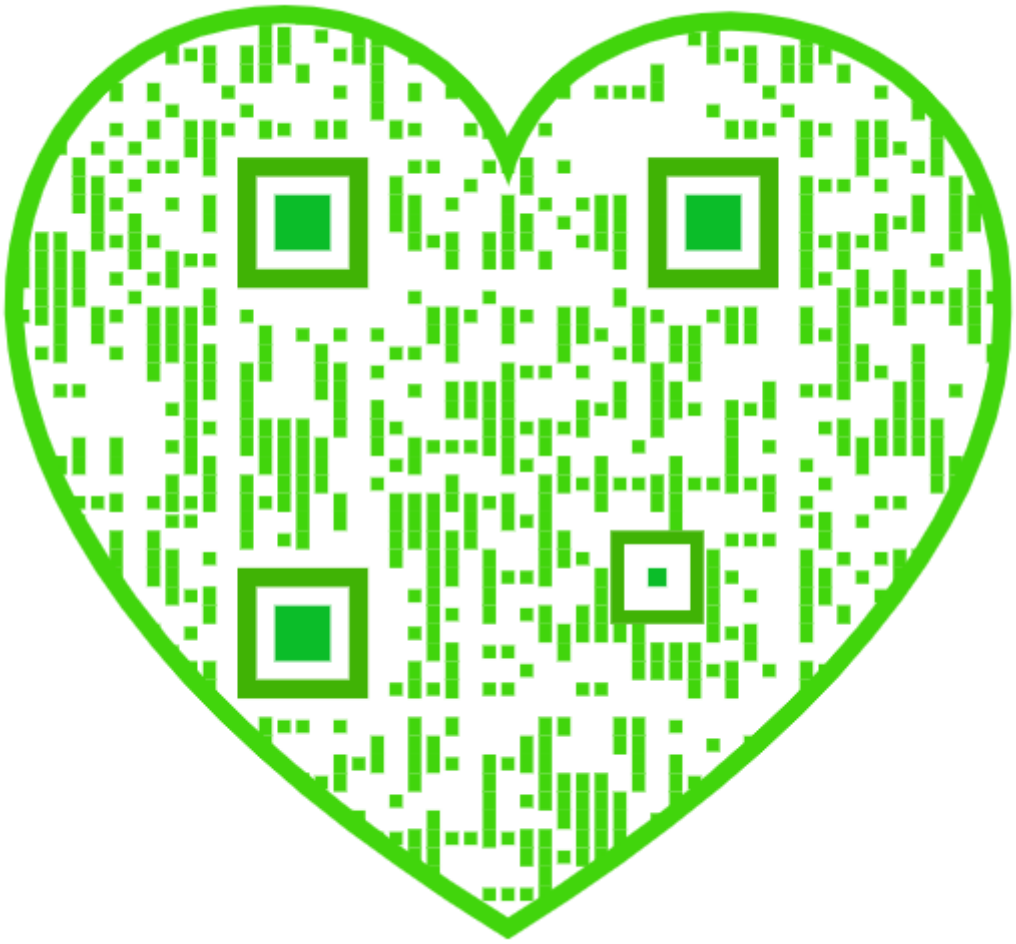
> "The hope is to give paralysed patients greater ability to interact with their environments."

Donoghue's team was supported by a private company in Massachusetts — Cyberkinetics Neurotechnology Systems — of which Donoghue is founder, director and chief scientific officer. Such hearty involvement by commercial interests is unusual, at least in this area of neuroscience. Experience from other scientific disciplines suggests that this route will lead to new challenges, particularly in the free and open dissemination of data (see *Nature* **442,** 1; 2006). However, these issues can be addressed, and they should not stop companies such as Cyberkinetics taking this knowledge from the laboratory to the clinic.

Both papers also owe a great deal to primate research, including the sort of curiosity-driven research most fiercely denounced by the animal-rights movement. Although some work can be done in rats, the best model of the human motor cortex is that of a monkey, the rhesus macaque. Scientists have spent years learning how the macaque's brain works, without necessarily planning to develop neuroprostheses. Experiments on primates will always be contentious, but this work shows why the immediate utility of a particular strand of research should not be the sole factor in determining whether it is ethical for it to proceed.

Although the videos of these latest experiments may garner a great deal of well-deserved attention, it is worth recalling that work in neuroprosthetics is by no means confined to spheres imprinted on the public's collective imagination by science fiction. Steve Austin was rebuilt to have enhanced strength, speed and vision; to be "better than he was before". The idea of giving people superhuman powers greatly appeals to the popular imagination. But in the real world, using neuroprosthetics to give patients control over all the less glamorous things we take for granted will be more important. ■

©2006 Nature Publishing Group



Building bridges

An American geneticist advocates a *rapprochement* with religion.

Public jousting between scientists and organized religion seems to have established itself as a consistent theme of the wider public discourse. Last week, it was the head of the Vatican's Pontifical Council for the Family letting it be known that researchers of the Catholic faith had better adhere to church doctrine on stem-cell research, or face possible expulsion. Next week it will be something else, and it will probably be in the United States, where relations between the two spheres have never been cordial and are becoming steadily less so.

To many scientists, religious contributions to public debates seem threatening and ill-considered. Religious leaders speak out against entire lines of enquiry — such as work on embryonic stem cells — in the name of God. They take stands against life-saving practices, such as condom use in areas of high HIV infection, in the name of morality.

Such contributions dismay the many scientists who are believers but who take a different doctrinal stance. They also irritate or enrage those (probably comparable in number) who are agnostics and atheists. After all, to many people, including scientists, the world simply makes more sense without the existence of God, and religious interventions are either offensive or irrelevant.

In response, some scientists are tempted either to publicly dismiss religious belief, or else to argue stridently against it. The latter approach is valuable in that it exposes religious dogmas to rational consideration and leads to their abandonment where they conflict with reality. But it is damaging if it fails to acknowledge the inability of science to deal with many of the issues that people face in their everyday lives.

An alternative response for believers and non-believers alike is to engage with people of faith to explore how science — both in its mode of thought and its results — is consistent with their religious beliefs. Many scientists have been quietly doing this for years. Their arguments are amplified by a new book from Francis Collins, director of the US National Human Genome Research Institute (see page 114).

In *The Language of God* (Free Press, 2006), Collins presents what he believes to be the counter-arguments to the doubts scientists have about God. He then expounds the rationale behind his own school of religious thought: theistic evolution, which posits that evolution is real, but that it was set in motion by God. He believes that it is the existence of God that explains certain aspects of humanity, such as self-sacrificing morality, as well as the fine-tuning of the fundamental constants that allow life to exist.

The book is unsparing in its criticism of both creationism and intelligent design. Both are false and unscientific, says Collins. He even argues that both ideas endanger religion itself, because they rest on such shaky ground that their backers risk losing credibility.

"Francis Collins presents what he believes to be the counter-arguments to the doubts scientists have about God."

Collins' style is straightforward, and even moving, especially when he discusses episodes from his own life, such as his difficulty coping with a sexual assault on his daughter. Even so, his reasons for believing in God and for becoming a devout Christian are unlikely to sway anyone who doesn't already believe.

But that's not the point. Collins is reaching out, from an exalted position in the world of science, to the realm of faith. By exploring, not least, how the Human Genome Project has added to our understanding of evolution, he hopes to provide a bridge across the social and intellectual divide that exists between most of US academia and the so-called heartlands, where religion is writ so large. Given the scale of the gulf, that is a laudable ambition. ■

The beautiful game

Punditry took a hiding in Germany.

One of the overriding messages from the World Cup that has just ended in Berlin is that football (that's soccer to our American readers) is almost impossible to predict. As a low-scoring game, it has an inherently stochastic quality that makes it gloriously exciting and palm-thumpingly frustrating in equal measure.

The struggle to anticipate World Cup results has taken many forms. In London, *The Guardian* newspaper attempted to verify the mantra of the Internet age that wisdom lies with the masses, inviting readers to vote for different betting options for each match. By the end of the tournament, the people made a profit, turning £250 (US$460) into £356. However, the newspaper's pet goldfish, which chose its bets by swimming to different parts of its tank, put them to shame, ending up with £369.

A look at more 'scientific' efforts at prediction turns up similar examples of painful hubris. A group of Norwegian mathematicians, for example, designed a computer model that simulated the complete tournament 2,000 times over (see www.nature.com/news). It predicted a Brazilian victory — but in reality, Brazil performed rather miserably and only made the quarter-finals.

Perhaps the tournament's least adroit piece of scientific punditry, however, came from Michael Shadlen, a neuroscientist at the University of Washington in Seattle. In an interview for *Nature*'s online World Cup preview, he hailed French maestro Zinedine Zidane as the world's most intelligent footballer. Zidane certainly made his mark, winning the Golden Ball award as the tournament's outstanding player — before being sent off in Sunday's final for a disgraceful headbutt on an opponent. Not too clever, really.

At least Italian scientists can take heart from that bombastic finale. The country's footballers have returned home in glory as deserved champions, to face a match-fixing scandal that could see several of the clubs that employ them relegated in ignominy. But as researchers there can testify, flourishing in the face of official incompetence and corruption is just what all Italian professionals have to do, every day of the week. ■

©2006 Nature Publishing Group

RESEARCH HIGHLIGHTS

CELL BIOLOGY

Central oasis in gene desert

J. Cell Biol. doi:10.1083/jcb.200603083 (2006)

Mammalian chromosomes contain gene-rich clusters scattered along stretches of barren, gene-poor 'deserts'. A recent study addresses how these seemingly desultory structures may be associated with the three-dimensional arrangement of chromosomes within the nucleus.

Together with their colleagues, Lindsay Shopland of the Jackson Laboratory in Bar Harbor, Maine, and Timothy O'Brien, now of Cornell University, Ithaca, New York, differentially labelled the gene-rich clusters and gene-poor deserts along a 4.3-megabase region of a mouse chromosome. They found that deserts frequently localized near the periphery of the nucleus, whereas clusters often occupied the centre.

This pattern of gene-rich and gene-poor clusters on this region of the chromosome is conserved though evolution from primates to chickens. This suggests that the patterns are in fact playing a role in positioning the various parts of the chromosome within the nucleus.

An eye for snakes

J. Hum. Evol. **51,** 1–35 (2006)

Why are children more frightened of rarely seen snakes than everyday hazards such as live electrical sockets and speeding cars? The answer might lie in the evolution of the primate visual system, speculates Lynne A. Isbell of the University of California, Davis.

Venomous snakes have a long, shared evolutionary history with primates, especially catarrhines — the Old World monkeys and apes, which includes humans. A proposed adaptation to be wary of snakes correlates with the expansion of those regions of the brain most associated with vigilance, learning, memory and fear, as well as binocular vision and a sugar-rich diet that allowed the brain to work faster. Whereas other mammals responded to snakes with physiological resistance to their venoms, catarrhines evolved to detect their enemy before the strike.

J. RINALDI/REUTERS/CORBIS

METEOROLOGY

Sky rivers

Geophys. Res. Lett. **33,** L13801 (2006)

Rivers in the sky can inundate rivers on the ground and cause them to burst their banks.

Martin Ralph of the National Oceanic and Atmospheric Administration in Boulder, Colorado, and his co-workers have found that all seven of the floods on the Russian River in northern California since 1997 coincided with the appearance of narrow streams of warm, humid air dubbed atmospheric rivers. For example, a moist corridor of air about 2,000 kilometres long flowing northeast into California from the Pacific Ocean, spotted in satellite microwave images, fed a flood-inducing storm in February 2004. Sometimes the atmospheric rivers triggered floods, sometimes they didn't — but floods were never triggered without one, say the researchers.

GEOPHYSICS

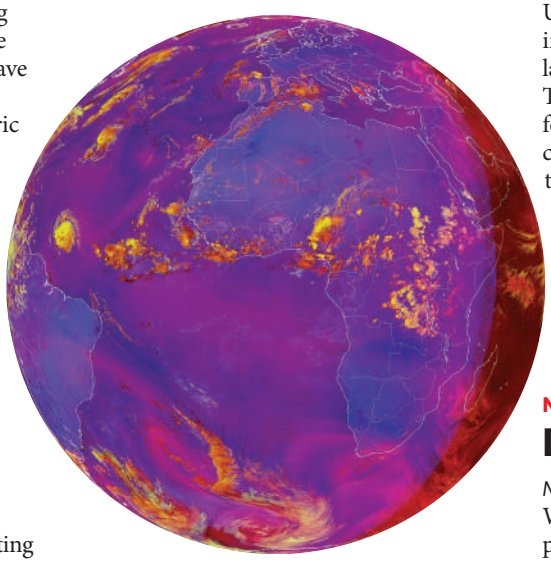
Earth wobble

Geophys. Res. Lett. **33,** L13303 (2006)

Tiny wobbles in the Earth's axis of rotation are caused by changes in weather patterns (pictured right), say Sébastien Lambert of the Royal Observatory of Belgium in Brussels and his colleagues.

Like a spinning top, the Earth's axis of rotation changes as it spins. Two cycles, lasting hundreds of days, dominate the movement of the pole. But when these cycles cancel each other out, smaller motions can be tracked.

Using Global Positioning System data, Lambert and his colleagues saw tiny changes, of less than a metre over four months, in the position of the pole. After examining weather data from the same period, the team concludes that the movement is driven by the appearance of regions of high and low atmospheric and oceanic pressure.



EUMETSAT

IMMUNOLOGY

Friend or foe

Nature Immunol. doi:10.1038/ni1362 (2006)

Telling an enemy soldier from an ally often relies on recognizing the uniform he wears. Now, researchers in Japan say the immune system does something similar to stop itself from attacking the 'friendly' bacteria that live in our guts.

A team led by Shizuo Akira of Osaka University showed in mice that a subset of immune cells found in the gut, known as lamina propria cells, express a receptor called TLR5. This recognizes the protein flagellin, found on the outer structures of disease-causing bacteria, and prompts the cells to trigger an immune response. But unlike some immune cells elsewhere in the body, lamina propria cells lack the TLR4 receptor, which recognizes the common bacterial molecule lipopolysaccharide. This combination lets them ignore helpful bacteria in the gut and target harmful ones.

NEUROBIOLOGY

Deadly dose causes Down's

Neuron **51,** 21–28; 29–42 (2006)

What causes the learning difficulties in people with Down's syndrome? Even though

©2006 Nature Publishing Group

there are hundreds of genes on the extra chromosome 21 that causes the condition, the answer could be surprisingly simple, say researchers.

Working in mice, a team led by Ahmad Salehi of Stanford University, California, has found that an overdose of just one protein made by a gene on chromosome 21 is enough to trigger the death of neurons important for cognition. The extra protein, APP, disrupts the function of another protein called NGF that controls how neurons develop connections with each other.

Meanwhile, Susan Dorsey and Lino Tessarollo of the National Cancer Institute in Frederick, Maryland, and their colleagues studied a receptor for an NGF-related protein called BDNF. The receptor comes in two versions, and increased levels of one kind led to increased neuronal death. Similar processes could be at work in neuro-degenerative conditions such as Alzheimer's disease, the teams say.

PLANETARY PHYSICS

Picture planets

Astrophys. J. (in the press); preprint at http://arxiv.org/abs/astro-ph/0601092 (2006)

Astronomers have come up with a method that they claim can sketch the surface of an extrasolar planet.

In a paper to appear in *Astrophysical Journal*, Peter Williams of the Harvard Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and his colleagues propose using the infrared Spitzer Space Telescope to image an eclipse of an extrasolar planet. Rather than watching the planet pass in front, the team will look for the loss of heat emitted by the planet as it goes behind the star.

The rate at which the brightness drops will vary if the surface of the planet is uneven, and with sufficient analysis, the authors claim it may be possible to see planetary features.

COSMOLOGY

Forwards to the past

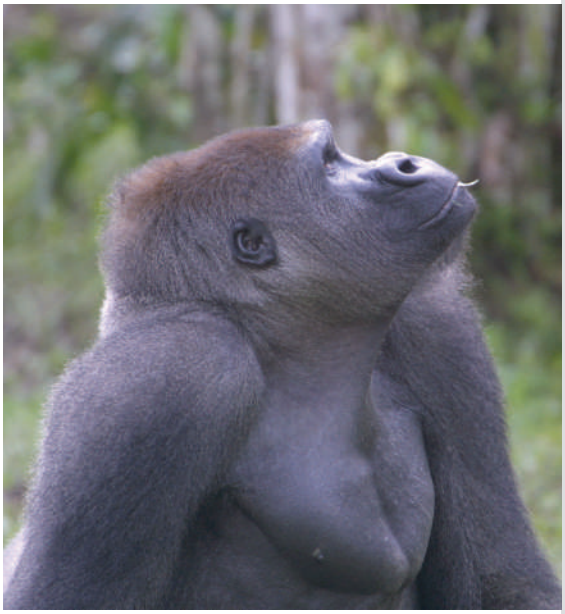
Phys. Rev. D **73,** 123527 (2006)

We're looking at how the Universe formed the wrong way around, say theoretical physicists.

Instead of trying to second-guess conditions at the start of the Universe and modelling how this affected its subsequent evolution, we should instead start with the Universe we observe today and work backwards, to see what initial conditions might have existed at its birth.

Stephen Hawking of Cambridge University, UK, and Thomas Hertog of CERN in Geneva, Switzerland, argue that we can't control the initial conditions of the Universe, or observe its evolution many times over.

Instead, scientists should base their calculations on scenarios whose beginnings are left unspecified, but that obey constraints in force later in the Universe's development. These constraints include the dimensionality and geometry of the Universe we observe today. Evolved backwards, such a scheme admits several alternative histories for the Universe.



D. CAILLAUD

CONSERVATION BIOLOGY

Ebola spread

Curr. Biol. **16,** R489–R491 (2006)

Ebola virus infections can spread through gorilla populations more easily than previously thought, according to researchers who have carried out the first study of an outbreak in the animals.

Since 2001, Damien Caillaud of Rennes and Montpellier Universities in France and his colleagues had been studying a population of about 400 western lowland gorillas (*Gorilla gorilla gorilla*) in Odzala-Kokoua National Park, Republic of Congo. Ebola struck in 2004.

Gorillas live in social units of family groups and solitary males, and the team found that the virus killed some 97% of group-dwelling individuals. But it also spread unexpectedly easily between social units, including solitary males — of which 77% were killed (such as Boniface, pictured above). Overall, 95% of the population disappeared in one year.

JOURNAL CLUB

Tarun Kapoor
The Rockefeller University, New York

Cell division has gained an extra dimension, says a cell biologist.

Seven summers back, I was at the Marine Biological Laboratory in Woods Hole, Massachusetts, studying the size and shape of the cell-division apparatus.

When a cell splits into two, its chromosomes are segregated by a structure called the mitotic spindle. This is made from microtubules, focussed together at each end of the spindle, which serve as tracks for the chromosomes to move along.

The components of the mitotic spindle are now mostly known, but it remains unclear how it self-organizes into a structure of the proper shape and size. One hypothesis invokes a second scaffolding — a non-microtubule 'spindle matrix'.

In the Cell Division Group at Woods Hole, we were using high-resolution microscopy to track the transport and polymerization of tubulin, the constituent of microtubules. It was difficult to explain our findings if microtubules alone determined the spindle size.

Inspired by these studies, I focussed on Eg5, a motor protein needed for the spindle to assume its bipolar shape. Imaging of Eg5 dynamics provided additional evidence for a non-microtubule spindle component. However, we were unable to obtain direct evidence for the spindle matrix, and my continued work on Eg5 provided more straightforward explanations for how it may function.

Since then, another team has proposed that a nuclear protein, lamin B, may be a component of the elusive spindle matrix (M. Y. Tsai *et al. Science* **311,** 1887–1893; 2006). The lamin B structures they describe recruit proteins important for spindle assembly, including Eg5. This adds a new dimension to the study of mitosis. Examining the properties of the lamin matrix and its interplay with microtubule-based structures will keep researchers busy for some time.

©2006 Nature Publishing Group

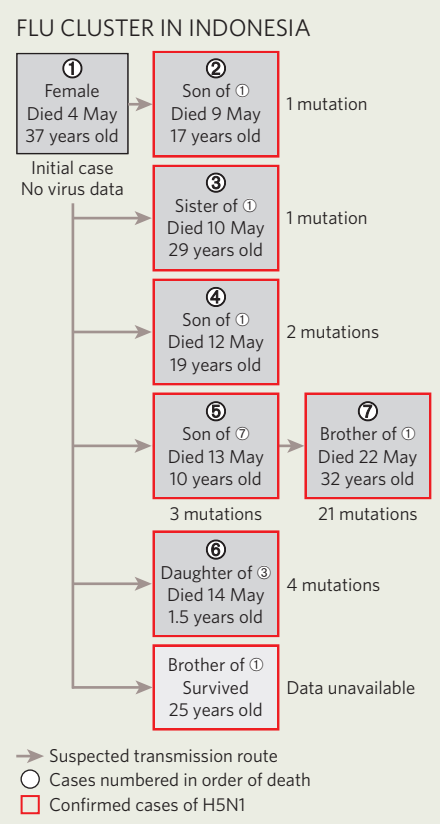
NEWS

Family tragedy spotlights flu mutations

A strain of avian flu that spread through a family in Indonesia, killing seven of the eight people infected, was accumulating mutations as it spread from person to person, according to confidential sequence data seen by Nature. The functional significance of the mutations isn't clear — most of them seem unimportant. But influenza researchers say the finding reiterates the need for sequence data to be made more widely available, if the virus is to be better understood.

The cluster of cases of the deadly H5N1 strain, which occurred earlier this year, is the first in which the World Health Organization (WHO) has admitted that human-to-human transmission was the most likely cause of spread (see *Nature* **441,** 554–555; 2006). Eight members of an extended family in Kubu Sembelang, in north Sumatra, were affected. The first patient, a 37-year-old woman who became ill on 24 April and died on 4 May, is thought to have caught the disease from poultry, then transmitted it to six relatives. One of these, her 10-year-old nephew, who died on 13 May, is thought to have passed the disease to his 32-year-old father (see 'Flu cluster in Indonesia').

Virus isolates from six of the eight family members have been sequenced, but the WHO has not released the data, saying that they belong to Indonesia. Instead, the agency released a statement on 23 May stating that there was "no evidence of genetic reassortment with human or pig influenza viruses and no evidence of significant mutations".



PARTLY SOURCED BY WORLD HEALTH ORG.

Nature has now obtained more detail on the genetic changes, which suggest that although the WHO statement was not incorrect, plenty more could have been said. Viruses from five of the cases had between one and four mutations each compared with the sequence shared by most of the strains. In the case of the father who is thought to have caught the virus from his son — a second-generation spread — there were twenty-one mutations across seven of the eight flu genes. This suggests that the virus was evolving rapidly as it spread from person to person.

One of the mutations confers resistance to the antiviral drug amantadine, a fact not mentioned in the WHO statement. The data were presented by Malik Pereis, a virologist at the University of Hong Kong, at a closed meeting of around a dozen international experts in animal and human health, held in Jakarta, Indonesia, in late June.

The virus did not evolve into a pandemic strain — the combination of mutations was not even enough to allow it to spread beyond close family. Many of the genetic changes did not result in the use of different amino acids by the virus. And there were no amino-acid changes in key receptor binding sites known to affect pathogenicity and transmissibility.

But experts say they cannot conclude that the changes aren't significant. "It is interesting that we saw all these mutations in viruses that had gone human-to-human," says one scientist who was present at the Jakarta meeting but did not wish to be named because he was commenting on confidential data. "But I don't think anyone knows enough about the H5N1

Genomics luminary weighs in on US faith debate

Is it really possible to combine dedication to science with belief in God? In a new book, prominent US scientist Francis Collins sets out his case for combining a strong religious faith with a zeal for the scientific method. But his views have already sparked debate, with critics suggesting that more talk of religion is the last thing that science needs.

Collins, who directs the National Human Genome Research Institute in Bethesda, Maryland, and headed the Human Genome Project, has never hidden the fact that he is a devout Christian. But he has never spoken quite so publicly about his faith. He says he felt compelled to write his book because the popular debate on faith and science has become dominated by extreme voices, leaving many feeling that there is no way to reconcile religious and scientific views of the world. "Our society is not well served by portraying a future which is either entirely secular or entirely religious in a fundamentalist way," he says.

Collins also hopes the book, *The Language of God* (Free Press, 2006), will provoke thought in academia, where, he says, the subject of faith isn't exactly popular. "In most academic circles, a discussion of spiritual matters tends to clear the room fairly quickly."

Discussion, Collins suggests, might rectify the misconception that most scientists are atheists. Surveys find that about 40% of US scientists believe in God, but Collins says that is not reflected in science's public face. That hurts science, he argues, because it drives away curious people who might also be religious believers.

"A few voices will be shouting out, saying 'wait a minute, this is nonsense'."

Collins takes a strong stand against some religious beliefs, such as creationism and 'intelligent design'. He considers both to be views that restrict faith to covering gaps in scientific knowledge, leaving it in a tenuous position.

Instead, Collins embraces a theology sometimes called theistic evolution, or BioLogos. This embraces the idea that human evolution occurred through natural selection according to God's plan, and that God instilled humanity with certain characteristics, including a 'moral law', that can't be explained by science.

©2006 Nature Publishing Group

VAN GOGH PAINTED PERFECT TURBULENCE
The disturbed artist intuited the deep forms of fluid flow.
www.nature.com/news

genome to say how significant that is."

Elodie Ghedin, a genome scientist at the University of Pittsburgh School of Medicine in Pennsylvania, says she's surprised that the virus from the father had so many mutations compared with others in the cluster, apparently arising in just a few days. "I have a hard time believing that the father acquired the virus from his son," she says, adding that the nine mutations in one gene in the father's virus are almost identical to those in viruses isolated from human cases in Thailand and Vietnam in 2004.

One possibility is that the father simply caught a different strain of virus from birds, although other mutations in his virus are similar to those in the strain isolated from his son. Or perhaps the virus from the son reassorted with another flu strain circulating in his father at the time, Ghedin says.

Part of the reason the picture is so unclear, say virologists contacted by *Nature*, is that the continued withholding of genetic data is hampering study of the virus. None of the sequence data from the Indonesian cluster has been deposited in public databases — access is restricted to a small network of researchers linked to the WHO and the US Centers for Disease Control and Prevention in Atlanta, Georgia.

Ghedin, for example, works on how mutations in one area of a genome can predispose other areas to further changes. She is part of a project started in 2004 to sequence thousands of human and bird-flu strains, but she has little access to H5N1 virus from humans. "Flu researchers don't all look at the data from the same angle," she says. "The more diverse analyses that are performed, the better we will understand the evolution of this virus."

"Flu researchers don't all look at data from the same angle."

"If all of the H5N1 isolates were available, there'd be quite a few people focused on understanding these data," agrees David Lipman, director of the US National Center for Biotechnology Information in Bethesda, Maryland.

But Paul Gully, who joined the WHO two months ago as senior adviser to Margaret Chan, head of the agency's pandemic-flu efforts, defends the agency's position. He points out that the WHO's priority is investigating outbreaks, not academic research. And he adds that although calls for more complete genome data and wider sharing of samples are "a valid point", labs are stretched during outbreaks, and don't have the time or resources to do high-quality sequencing.

He agrees that sharing samples with other researchers would allow such work to be done. But he says the WHO must work within the constraints set by its member states — they own the data, and decide whether to share it. "As more countries share data, hopefully that research will get done," he says.

The WHO has not formally asked Indonesia to share the sequences, Gully adds. "We would rather wait and see what Indonesia decides." ■

Declan Butler

AP/BINSAR BAKKARA

Johannes Ginting is the only survivor of a worrying cluster of bird-flu cases in Indonesia.

AP/D. C. PIZAC

Francis Collins is a devout Christian.

"The moral law is a signpost to a God who cares about us as individuals," Collins says. "God used a mechanism of evolution to create human beings with whom he could have that kind of fellowship."

Many scientists disagree strongly with such arguments. Some suggest that science is on the defensive today — not just in the United States — and that society needs exactly the opposite of what Collins suggests: less talk about faith and more about reason. Religious concerns are largely behind the US law restricting federal funding of stem-cell research, for example. And many feel threatened by the influence of intelligent design in science education.

In the United States, "the default position right now is to assume that religion is perfectly OK", says Paul Myers, a biologist at the University of Minnesota in Morris and author of the popular science blog Pharyngula. "Collins is taking that default position, and while a large majority of scientists will shrug their shoulders, a few voices will be shouting out, saying 'wait a minute, this is nonsense'."

"I cannot see how this could be good for science — supernaturalism is fundamentally anti-scientific," says Richard Dawkins, a biologist from the University of Oxford, UK. "Scientists work hard at trying to understand. Supernaturalism is an evasion of this responsibility. It's a shrug of the shoulders."

Dawkins acknowledges that, particularly in the United States, there might be tactical reasons for trying to get on with religious people. "That is a perfectly reasonable political stance, but it has nothing to do with truth."

Others welcome Collins's book, however. "I think it's helpful when scientists of Francis's prominence speak out on the compatibility of faith and science," says Eugenie Scott, executive director of the National Center for Science Education, a group based in Oakland, California, that lobbies against creationism.

Scott agrees with Collins that so far the harshest voices have achieved most prominence, and that this situation doesn't help either side. "Creationists love quoting Dawkins and Daniel Dennett," she says. "But those individuals don't represent the fairly sizeable proportion of non-theists who are not out to destroy religion." ■

Erika Check

©2006 Nature Publishing Group

SNAPSHOT

DEEP-SEA WONDERS

From the whimsical to the downright scary, images featuring creatures from the deep are showcased in the BP Kongsberg Underwater Image Competition being held this week at the 11th International Deep-Sea Biology Symposium in Southampton, UK. **Narelle Towie** takes a look at some of the most striking entries.

D. FORCUCCI WWW.UNDERSEAIMAGES.COM

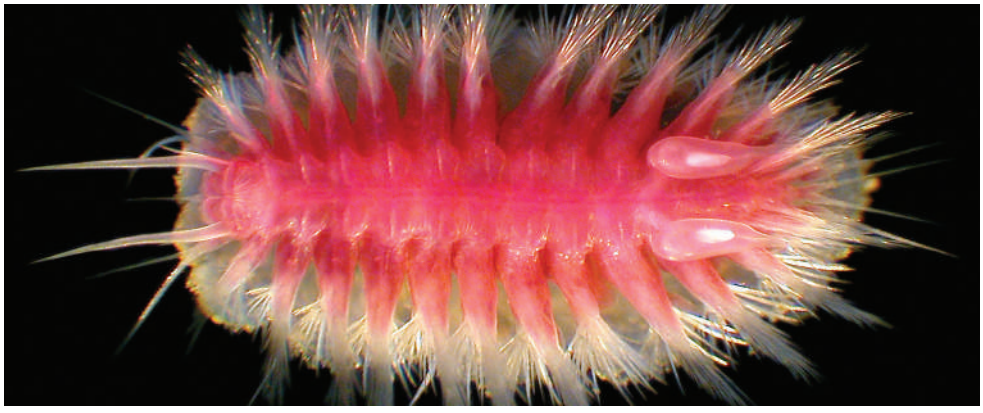
Looking like a set of deep-sea headlights, the luminous stare of this myctophid lanternfish was captured by David Forcucci as he worked aboard a research ship in pirate-infested waters off the Somali coast. Caught in a net at 2,000 metres depth, the fish was one of many hauled aboard the RV *Malcolm Baldrige* in 1995 as part of the Globec programme, which studied marine populations in the Indian Ocean. The image earned Forcucci fifth place in his category.

During the day, lanternfish lurk in gloomy waters hundreds of metres deep to avoid predators. But as sundown approaches, the fish rise much closer to the surface to feed on zooplankton.

Six years after this picture was taken, a US oceanography research ship working in the same area was chased and shot at by pirates (see *Nature* **413,** 97; 2001).

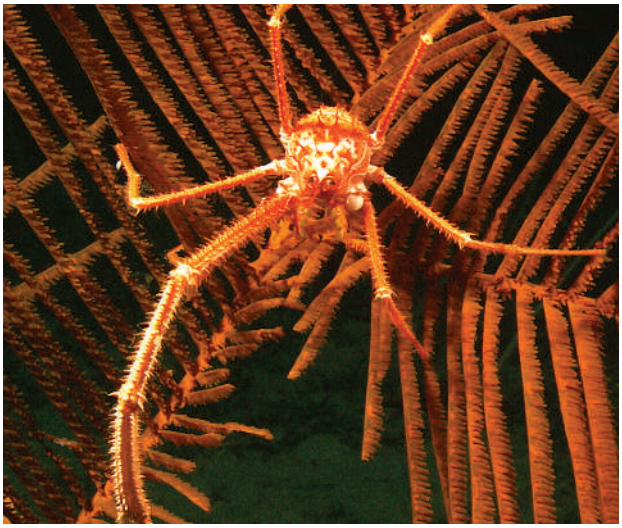
A. GLOVER, NAT. HIST. MUS., LONDON

"It's a snowboarding whale worm," says photographer Adrian Glover, a researcher from the Natural History Museum in London. "Not really, but that's what *Bathykurila guaymasensis* looked like when we found him," he explains. This polychaete, collected by researchers studying whale falls in the Santa Cruz Basin, southern California, was found 1,600 metres down, moving across a white mat of bacteria that were feeding off the bones of a dead grey whale.



©2006 Nature Publishing Group

S. C. FRANCE, UNIV. LOUISIANA, LAFAYETTE



This frame grab from a video camera attached to a remotely operated vehicle was bagged by Scott France from the Deep Atlantic Stepping Stones research group, and came in the last minutes of a 19-hour dive to study coral communities in the Mid-Atlantic Ridge.

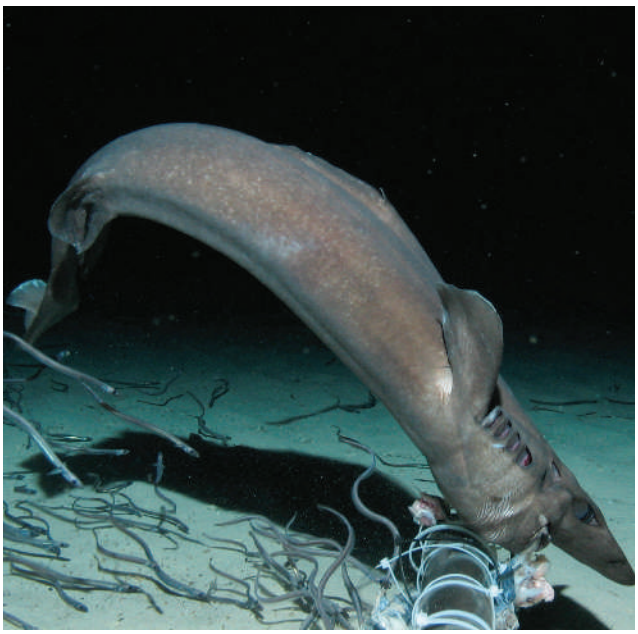
Perched high on a coral, this five-armed galatheid crab was lucky to have a seat at all. The 2-metre-tall rare black coral shown here narrowly missed being wiped out by the trawl of a fishing net that destroyed coral directly adjacent to this shot. The image was taken in August 2005 and won its category.

D. SHALE

The vivid orange colours of this *Stauroteuthis syrtensis* were caught by wildlife photographer David Shale during a research cruise in the Mid-Atlantic Ridge. The image won its category.

Nicknamed Dumbo because of two small flipper fins behind its eyes, this 30-centimetre-long octopus was snared in a collection tank attached to the *Johnson-Sea-Link*, a small submersible capable of diving to 1,000 metres. Both human occupants and sea life are kept at a chilly 6 °C while in the submersible. Brrr.

N. KING, OCEANLAB, UNIV. ABERDEEN



This deep-sea demersal shark (*Deania calceus*) fed for 20 minutes before letting hungry arrowtooth eels have their turn on the metal feeding arm of the RObust BIOdiversity lander — known as ROBIO. By handing out mackerel to scavenging fauna, ROBIO gives researchers from Oceanlab at the University of Aberdeen, UK, the ability to survey sea life.

This shot, taken at 931 metres depth in sea waters west of Ireland, earned a 'highly commended' rating for Aberdeen's Nicola King.

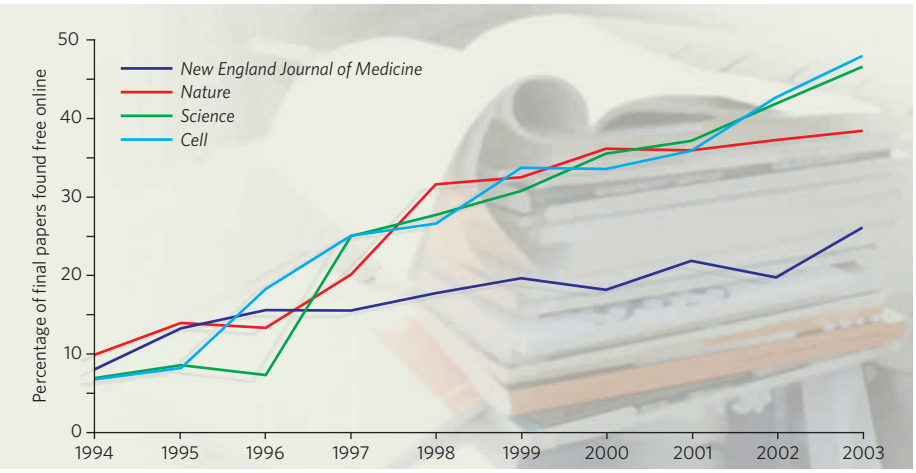
©2006 Nature Publishing Group

PS I want all the rights

The debate over whether research results should be freely accessible has always been fraught. Having given a lot of ground, journal publishers are determined to hang on to one last bastion: their rights to the published version of scientific articles. Now librarians and open-access advocates have set their sights on that final prize — by encouraging researchers to demand the right to distribute the published versions freely and immediately.

Funding agencies are increasingly adopting policies to make the results of the research they fund free for all. Both the US National Institutes of Health (NIH) and Britain's Wellcome Trust, for example, encourage this practice. They ask that the version of a manuscript accepted for publication be put in an open-access library such as PubMed Central within 6–12 months of it coming out. The Wellcome Trust's policy will become compulsory on 1 October, and legislation that would make the NIH policy mandatory is pending in the US Congress (see *Nature* **441,** 915; 2006).

Ann Wolpert, director of libraries at the Massachusetts Institute of Technology (MIT) in Cambridge, Massachusetts, has launched an initiative that she says will clearly assign rights to the author in a way that would satisfy funders. Wolpert has drawn up a document that researchers can add on to the rights agreement the publisher gives them to sign. Similar agreements have been drafted by the Scholarly Publishing and Academic Resources Coalition and the MIT-affiliated Science Commons.



SOURCE: J. D. WREN *BR. MED. J.* **330,** 1128–1131 (2005)

More than a third of authors already post final versions of their science papers online, a *BMJ* study found.

"I look at it as responding to a request by faculty members to simplify their lives," says Wolpert. "They say 'it is crazy that we are supposed to read and understand these publishers' agreements. Give me something that I can just staple to any agreement, so I can comply with NIH or Wellcome Trust policy'."

But publishers' groups argue that the agreements being drafted go much further than is necessary to comply with current policies. Wolpert's document, for example, would allow authors to publish the final, formatted version of their paper anywhere on the Internet, as many times as they like, immediately after publication. "This isn't a balance of rights. This is giving MIT all the relevant rights," says David Hoole, head of brand marketing at Nature Publishing Group.

Publishers point out that most journals already allow authors to post the accepted version of their paper online, as required by the NIH and the Wellcome Trust. Such versions have been peer reviewed, but aren't copy-edited, formatted or paginated. But giving authors

City state hopes research cash will buy global status

Singapore already has a reputation for throwing money at scientists in particular fields it wants to develop. Now, the government says it will almost double its research budget over the next five years, to a hefty 3% of its gross domestic product (GDP). But whereas those in the favoured fields have welcomed the money, others complain that, as ever, basic research is losing out.

The government announced on 7 July that its newly established National Research Foundation (NRF) will spend S$1.4 billion (US$876 million) during 2006–10 in three areas: biomedical research; environment and water technologies; and interactive and digital media. This is part of the S$5-billion five-year budget set aside for the NRF when it was created in January.

Another S$1 billion is earmarked for a project to attract world-class research institutes. As a start, the NRF says it will establish a joint research centre with the Massachusetts Institute of Technology next year in Singapore, and hire 300–400 researchers to work there. How the remaining S$2.6 billion will be spent has yet to be decided.

The announcement represents the first big move in the government's ambitious plan to increase its research and development budget to S$13 billion for 2006–10, or 3% of the country's GDP.

Singapore is famous for big science initiatives, especially in the biomedical sector. One example is Biopolis, a S$500-million development that houses biotech research institutes and pharmaceutical giants such as Novartis. Next year, the government plans to open the first phase of a Fusionopolis complex, which will house information-technology and media companies.

Despite previous investment, the NRF's chairman, Tony Tan, says the Singaporean government is concerned about keeping its competitive edge in science and technology research, compared with big and growing economies such as those of India and China.

Over the past few years, Tan has travelled to other small but wealthy countries such as Switzerland, Denmark and Sweden to find out how they do it. He says he learned the importance of specializing in a small number of research areas that are likely to turn a profit.

The NRF is now discussing exactly how funds in each of the three research areas will be spent and when to start accepting grant applications. S$550 million will go towards biomedical research, especially translational and clinical studies — a move from the previous

©2006 Nature Publishing Group

rights to the final versions, they say, could make it impossible for journals to earn a living.

Jerry Cowhig, managing director of the publishing arm of the Institute of Physics, says that the institute provides articles free online for 30 days after publication, and that he is happy for authors to post the accepted versions of their papers. But he is not in favour of making the final, edited version of a paper freely available everywhere. "That would be a real threat to the continuation of established journals, and the eventual outcome would be to damage scholarship," he says.

Sally Morris, chief executive of the Association of Learned and Professional Society Publishers, agrees. "The final version is where publishers add value," she says.

On 27 June, Morris's group, along with the International Association of Scientific, Technical and Medical Publishers, wrote to Wolpert outlining their concerns and proposing a meeting. A similar letter went out from the professional and scholarly publishing division of the Association of American Publishers on 7 July.

"This isn't a balance of rights. This is giving MIT all the relevant rights."

But would giving authors such rights really damage journals? After all, many authors already post the final version of their paper on the web regardless of what their rights agreement says. A study by Jonathan Wren, a bioinformatician at the University of Oklahoma in Norman (J. D. Wren *Br. Med. J.* **330,** 1128–1131; 2005), revealed that the final versions of more than one-third of articles in high-impact journals are freely available online (see graphic).

Ted Bergstrom, an economist at the University of California, Santa Barbara, argues that for libraries and other users, the convenience and authority of journal subscriptions will still be preferable to searching out free versions of papers individually. So agreements such as Wolpert's shouldn't affect the bottom line of any but the most overpriced publications, he says.

Wolpert adds that the value of journals isn't just in locating and reading individual papers, but in browsing and the 'serendipity factor'. "There is value behind a collection of articles judged worth your attention by smart people in your discipline," she says.

John Cox, a consultant to publishers and academic societies who is based in Chichester, UK, says that the value of papers as they appear on journal websites is often underestimated. "It's not just the copy-editing, but the infrastructure that is provided: the linking to citations, indexing, alerting services, the presentation of the product on the screen to the reader."

But he argues that the desire to post final versions across the web is misguided because the version published on the journal website will always be definitive. "It becomes, if you like, part of the minutes of science," he says. "That is deeply embedded in the scientific research culture." ■

Emma Marris

R. RAHMAN/AFP/GETTY

Will investment in research help Singapore keep its competitive edge?

focus on basic research. S$330 million is likely to go towards strengthening technologies that allow the country to produce clean water, for domestic use and for export. S$500 million will go towards developing media technologies such as video games and digital cinema.

Some academics, such as Barry Halliwell, deputy president of research and technology at the National University of Singapore (NUS), have acclaimed the funding, saying it will help to attract big names from overseas.

Others are underwhelmed. The NUS's Juan Walford, who studies seahorses as a barometer for the quality of the marine environment, says the new categories are just a reclassification of areas already given the bulk of government money. "There is really nothing new," he says. "It's all applied technology and engineering, there's no new opening for 'scientific research'." ■

Ichiko Fuyuno and David Cyranoski

SIDELINES

ON THE RECORD

"I'll have time for feelings after I'm dead. Right now, we're busy."

NASA administrator Michael Griffin tells reporters how he feels about the successful launch of the space shuttle Discovery on 4 July.

"I am really, really proud. The only problem is that I really don't know what to do with it."

Stefan Trellenkamp of the University of Kaiserslautern explains how he engraved the world's smallest soccer pitch — 500 nanometres long — on acrylic glass.

Sources: NPR, Reuters

SCORECARD

▲ **Mushrooms**
Researchers show that under carefully controlled conditions, hallucinogenic mushrooms can cause spiritual experiences that have positive effects on a person.

▼ **Cannabis**
Neuroscientists reveal that rats given cannabis when they are young are more likely to become addicted to heroin. This suggests that marijuana could make users more susceptible to hard drugs.

▼ **Obesity**
A US study of more than 9,000 adults finds that those suffering from obesity are 25% more likely to have mental health problems.

NUMBER CRUNCH

Scientists are drowning in paperwork, according to a survey of more than 6,000 US faculty members by the Federal Demonstration Partnership.

42% of 'research time' is actually spent doing administration.

4 hours a week could be saved if researchers got administrative help.

10% of their research grant is what some scientists say they would be prepared to pay for that help.

Source: Chronicle of Higher Education

©2006 Nature Publishing Group

Is India's 'patent factory' squandering funds?

NEW DELHI

'Patent or perish' is the slogan of Ragunath Mashelkar, head of India's largest publicly funded scientific agency. Over the past decade he has turned the 40 or so labs under his control into a patent factory. "Our labs obtain more patents in the United States than all Indian inventors combined," boasts Mashelkar, who directs the Council of Scientific and Industrial Research (CSIR). "To be noticed, you need a portfolio of patents."

But not everyone is impressed. Critics are accusing the council of wasting taxpayers' money by patenting hundreds of spurious inventions that are never exploited.

The approach of the CSIR has been to file a patent on any new finding that meets the criteria, whether or not the agency wants to commercialize it, says R. K. Gupta, the CSIR's chief of patents. That strategy has certainly been successful in terms of quantity. Between 2002 and 2006, the CSIR was granted 542 US patents — more than the total number granted to its counterparts in France, Japan and Germany combined.

But according to Suresh Chandran, who handled biotech patents in India before becoming licensing manager for ES Cell International in Singapore, "Most of the patents are not even worth the paper they are printed on." He points out that each US patent costs the CSIR US$25,000 for filing and $4,000 a year for maintenance. "Maybe it is a passing phase, but I can tell you that we waste quite a lot on this activity," he says.

One example given by critics is a cow urine extract that the CSIR patented in the United States in 2002, claiming that it enhanced the activity of antibiotics. The claim is yet to be backed by a peer-reviewed publication and the CSIR admits that no drug company has shown interest. "We had a good laugh when we read about it, and that was it," says a spokesman for a leading drug company in New Delhi, who asked not to be named.

Masheklar disputes that the patents are a waste of money, pointing out that a cluster of three US patents on a potential anticancer molecule has been licensed out to an Indian entrepreneur in the United States for around $100,000. Considering that only about 3% of US patents are ever licensed, it is too early for the CSIR to expect big returns, he says.

"A lot of money is going down the drain."

Critics counter that although the strategy of patenting everything has created awareness among scientists of the potential worth of their discoveries, it is time to ensure that patents create products and wealth, not just statistics. A. V. Rama Rao, former director of the CSIR Indian Institute of Chemical Technology in Hyderabad, wants the CSIR to set up an independent division to decide which developments are worth patent applications. "A lot of money is going down the drain in the name of patents," he says. ■

K. S. Jayaraman

©2006 Nature Publishing Group

Misconduct scandal delays Japan's research funding

Japanese research funding has been put on hold following a misconduct case involving Kazuko Matsumoto, a chemist at Waseda University in Tokyo.

Partly in response to the scandal, Japan's finance ministry has delayed distributing ¥10.6 billion (US$90 million) in grants to about 50 universities and research institutes. It has instead requested that they submit more detailed breakdowns on how they plan to use the money.

The decision comes after Matsumoto acknowledged misappropriating more than ¥14 million in research grants over the past few years. Late last month she offered to resign her university post, and has stepped down from her vice-presidency of the International Union of Pure and Applied Chemistry.

Waseda University is investigating further allegations that Matsumoto misappropriated more than ¥20 million and fabricated data.

Hurricane jet grounded over tussle with union

Some 'hurricane hunter' research flights have been grounded for being too dangerous. A US federal judge ruled on 30 June that the National Weather Service cannot send its Gulfstream jets to take measurements in a densely clouded region above the inner core of tropical storms. Flights may resume if the agency negotiates safe flying conditions with the labour union representing civilian members of the flight crews based in Tampa, Florida.

The decision applies only to the Gulfstream-IV jet, a lightweight plane that can fly at high altitudes. The Gulfstream is more susceptible to turbulence than the heavier WP-3D Orion, which is also used for storm research.

A National Weather Service spokesman, Jordan St. John, says that no plans have yet been made to negotiate with the union. But he adds that the agency still plans to equip the Gulfstream with an advanced Doppler radar system over the next few years, in the hope of flying it on the fringes of storms.

NOAA

Flying into a storm: the National Weather Service's Gulfstream (front) is grounded in a safety row.

National Academy sees red over Mars missions

NASA should change its plans for exploring Mars in the coming decade, says a panel convened by the US National Academy of Sciences.

The agency's current vision "is not optimized" for maximizing the scientific output of future missions, the panel concluded in a 6 July report. It recommended bringing forward the launch date of a long-term sensor network to monitor the martian structure and weather. But it suggested delaying the Astrobiology Field Laboratory until 2018 to allow "an informed decision of its merits" as well as to consider how findings from the Mars Science Laboratory (pictured) might affect its mission.

NASA

The panel also called on NASA to step up technological development for future Mars missions, particularly one to bring back rocks from the red planet.

ESA astronaut set to boost science on space station

ESA

Thomas Reiter: ESA's man in space.

European Space Agency (ESA) astronaut Thomas Reiter arrived last week on the International Space Station. This takes the station crew up to three permanent members for the first time since May 2003, after the space shuttle Columbia disaster.

The German's arrival, on the space shuttle Discovery, should free up the crew to spend more time doing science. During the next five to seven months, as well as helping to maintain the space station, Reiter will participate in about 20 experiments for ESA's Astrolab mission, ranging from plasma physics to investigating the effects of weightlessness on human physiology.

Court settlement puts sea lions back under scrutiny

US researchers headed into the Pacific Ocean last week to begin studies of threatened Steller sea lions — the first that have been permitted since a court shut down their projects.

All Steller research was blocked in May after a 2005 Humane Society lawsuit alleged that the species could be harmed by the research procedures used by the US National Marine Fisheries Service (NMFS) (see *Nature* 441, 677; 2006).

An agreement between the Humane Society of the United States and the NMFS was approved in court on 30 June and allows for low-impact studies.

As part of the agreement, research methods such as hot branding, tooth extractions, and researchers entering rookeries will not be permitted. For now, air surveys and observational studies will be allowed. A comprehensive environmental analysis will also be conducted to determine the effects of recent studies on the mammals.

Hwang admits being responsible for faked data

In a court in South Korea last week, Woo Suk Hwang admitted to having overall responsibility for two fraudulent papers on human therapeutic cloning published in 2004 and 2005 in *Science*.

Hwang was indicted in May on three charges: embezzlement, for misusing government funds given for research; fraud, for using knowingly falsified data to secure funding from private companies; and breaking a bioethics law by paying for human eggs. He is pleading innocent on all counts.

Hwang admits to ordering the fabrication of some data, but he says he believed that the stem cells were real and that he was merely deceived by a junior researcher. He also claims that the cell line presented in 2004 was a clone, despite a Seoul National University investigation that determined it was not.

His lawyer, Geon Haeng Lee, explained Hwang's expenditures by saying that confusion arose after Hwang pooled his own personal money with research funds.

Lee announced two weeks ago that Hwang will begin research again this month with a group of about 30 researchers and private funding. Hwang would work with animal cells, having lost his licence to work with human subjects.

©2006 Nature Publishing Group

BUSINESS

Giving it away

Charities are starting to operate like venture capitalists, putting their cash into fledgling drug companies. **Virginia Gewin** reports.

Robert Beall, a 63-year-old biochemist turned patient advocate, believes it is past time for charities to change their approach to finding treatments and cures.

"By absorbing some of the early financial risk involved in developing new drugs, philanthropists can increase the odds that a company will take on a small disease such as cystic fibrosis," says Beall, who heads the Cystic Fibrosis Foundation (CFF), a medical charity based in Bethesda, Maryland, with assets totalling US$230 million.

In 2000, the foundation established a non-profit body, CFF Therapeutics, to conduct clinical trials, maintain thorough patient records, and work on drug discovery and development for the lung disease. In 2003, for example, CFF Therapeutics spent US$1.7 million on early-stage clinical trials of denufosol, an early intervention treatment that is going into large-scale, phase III trials with Inspire Pharmaceuticals of Durham, North Carolina.

It is an approach that is gathering steam among philanthropic organizations, who are no longer content to attack disease solely through the well-established channel of backing biomedical research projects. Most want cures at once, if not sooner — and in their efforts to speed things up, they are acting more and more like venture capitalists. "Hundreds of millions of dollars, possibly billions, are being directed at biomedical venture philanthropy," says Greg Simon, president of FasterCures, a think-tank in Washington DC that is devoted to accelerating the medical-research process.

The real venture capitalists, meanwhile, are becoming a touch more risk-averse. Many of them are opting to invest funds in fewer companies at later stages of development. This climate encourages the neglect of possible treatments for thousands of 'orphan' diseases such as cystic fibrosis, whose small scale results in limited market potential.

Charities set up to help people with such diseases are starting to fill the gap, looking for roles in the early stages of developing treatments. "Disease organizations realize they can make a difference by identifying opportunities for industry," says Warren Lammert, director of the Epilepsy Therapy Development Project (ETDP) in Reston, Virginia.

Lammert is also a founder of Granite Point Capital, an investment firm based in Boston, and he has applied this expertise to the epilepsy project, which supports research as well as investing small amounts of money in young companies with treatment ideas. Since 2000, the ETDP has made five such investments — three of which have gone on to raise significant amounts of commercial venture capital.

Exporting expertise

Like venture capitalists, the new venture philanthropists try to provide their small companies with top-rate management support. The Institute for the Study of Aging (ISOA) — a centre in New York set up by the Estée Lauder family — has done this since 2000 for 16 companies investigating possible treatments for Alzheimer's disease .

Allon Therapeutics of Vancouver in Canada is one such company. It was started in 2001, in part with US$250,000 from the institute. The charity also helped co-founder Illana Gozes, a biochemist at Israel's Tel Aviv University, to write the first business plan, to license the intellectual property jointly held by Tel Aviv University and the US National Institutes of Health, and to appoint its first chief executive. The company attracted venture capital, went public on the Toronto Stock Exchange in 2005 and is currently valued at Can$33 million (US$30 million).

"Venture-capital investors value the fact that we've vetted a company scientifically," says Howard Fillit, chief executive of the ISOA. Much like a university, the ISOA says that it seeks a return on its investment for assuming the initial risk — but it is a modest one, which is funnelled back into other such ventures.

"Philanthropists know what's going on in the lab, understand disease, and realize what matters to patients."

These venture philanthropists want to make money as well as pushing along ideas for cures. One such outfit is Accelerate Brain Cancer Cure (ABC^2), a non-profit funder of research and entrepreneurial efforts founded in Washington DC by Dan Case and his brother Steve, former chairman of AOL, after Dan was diagnosed with a brain tumour. It has worked with other brain-disease charities to fund research on common hurdles to treatment — such as getting drugs to their target. In the past year, ABC^2 has set up the Brain Trust Accelerator Fund, to support young companies. It says the fund has three objectives: a return for investors, advancing therapies for brain diseases and contributing a portion of the proceeds to brain-disease charities.

Six-year-old Tanner Buck will be trying a drug for cystic fibrosis that was developed by a charity.

The mix of goals raises some grey areas, of course: investors must agree that a fund will pursue therapies, not just profits. But Lammert points out that cooperation with for-profit investment funds can get charities access to far more capital than they would otherwise attract.

Charitable status

Venture capitalists are keenly aware of the charities' interest. "Early stage investing is morphing," says Nicola Campbell of Sofinnova Ventures in San Francisco, which backed one of Lammert's investments: Marinus Pharmaceuticals in Connecticut. She says that her company is working with philanthropic bodies because they know what is going on in laboratories, have a good understanding of disease, and realize what matters to patients. She adds that the charities are realistic and candid about risks — a big plus in an industry that often gets only the rosy view of an idea from its inventors or company founders.

Indeed, philanthropy's greatest contribution

©2006 Nature Publishing Group

S. STREBLE

is its long-term perspective, says Paul Antony, chief medical officer of the drug-industry organization PhRMA. "There are a growing number of non-profits working on disease treatments, although they have different levels of management experience and scientific sophistication."

But failure to meet heightened expectations — as well as any appearance of profiteering — could deter donations to the charities themselves. "My fear is that philanthropists get seduced into making claims and promises they can't keep," says Frank Douglas, director of the Center for Biomedical Innovation at the Massachusetts Institute of Technology.

Christine Copple, chief executive of Starise Ventures, a consultancy based in Potomac, Maryland, says she hopes to attract private capital back to early stage deals by writing agreements that can satisfy the cultures of both charity and venture capitalism. Their shared goal — accelerating cures — fulfils philanthropy's mission and helps investors to create value, she says.

Only time will tell us how to balance cures and profits to donor and investor satisfaction. Neither the CFF nor the ETDF discount the potential of their own venture funds. But in the meantime, Beall says, the CFF will drive its own destiny and defend its ability to focus solely on patient needs. "Our goal is to create drugs, not profit for individuals," he says. ■

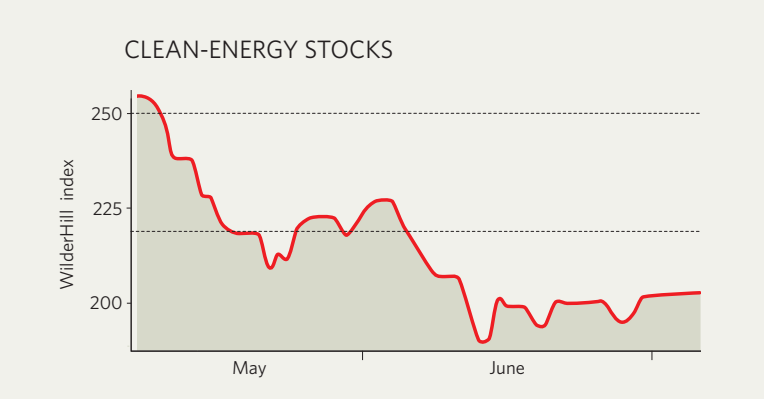
IN BRIEF

END OF THE LINE The German biotechnology company whose TGN1412 drug caused severe adverse reactions in six London volunteers in March has filed for bankruptcy. A statement from Würzburg-based TeGenero said that it was going into receivership because it would be impossible to secure the investment necessary to keep the 15-person company going. Lawyers for the volunteers — who face amputations and other severe after-effects — expressed concern that the bankruptcy would limit their prospects for compensation (see *Nature* 440, 388; 2006).

BLIND AMBITION California biotechnology overlord Genentech has won approval for Lucentis, a drug to treat a common loss of vision known as wet age-related macular degeneration. Although not unexpected, positive sentiment surrounding the US Food and Drug Administration's decision on 30 June sent the company's stock soaring: on 6 July, it closed at almost US$86, up from $79 a week earlier. Investor excitement was fuelled by the company's announcement that it will sell a single injection of the treatment ranibizumab for $1,950: thousands of elderly sufferers are expected to need it five or six times a year.

CHIPS IN Global semiconductor sales are surging forward on the back of demand for chips used in mobile phones and other portable devices, the industry reports. Total sales in May were $19.7 billion, up 9.4% on a year earlier, says the Semiconductor Industry Association — despite a small dip in the value of chips sold for use in personal computers. Demand for analogue chips for mobile phones grew by 21%, and total chip sales in Asia, excluding Japan, went up 15%.

MARKET WATCH



SOURCE: WILDERSHARES

Fears that stocks in clean-energy firms were becoming overvalued were expressed in this column two months ago (see *Nature* **441,** 281; 2006). Sure enough, they've fallen back to Earth with a bang.

The drop is sharper than the sector would have wanted, and has been driven not just by the earlier hype for clean energy, but also by the broader retreat from risk that characterized May's stock-market correction.

The WilderHill Clean Energy Index (ECO on the American Stock Exchange) measures the performance of renewable-energy companies listed in the United States — but its fall reflects a global trend.

Michael Liebreich, chief executive of New Energy Finance in London, says that aside from the general market downturn, three factors made stocks tumble in May: profit-taking; bad vibes from the collapse of the European carbon-emissions market (see *Nature* **441,** 405; 2006); and the plain fact that some stocks were overpriced. "A little bit of exuberance has been taken away," he says, "and most people in the industry are quite pleased about that."

A series of successful public share offerings — culminating in a billion-dollar haul for Norwegian solar-power company REC on 9 May — had earlier seemed to confirm clean energy's arrival as the next big thing.

Liebreich thinks that subsequent setbacks should serve to remind investors that there are constraints on how fast the industry can grow, such as the numbers of trained engineers, and the size of manufacturing facilities.

"There's a bit of worry that as policy-makers continue to plough in support, the industry will not be able to respond quickly enough," he says. "There's a limit to the rate of growth, and maybe we're coming up against it." ■

Colin Macilwain

©2006 Nature Publishing Group

R. MASSEY/GETTY

IN SEARCH OF THE SIXTH SENSE

Implants in the brain could one day help paralysed people move robotic arms and legs. But first, scientists need to work out how our brains know where our limbs are, says **Alison Abbott**.

Shut your eyes. Now, touch your nose. Chances are you can do this without even thinking about it. For this you can thank your sense of proprioception, which is so much a part of us that most of us are unaware that it exists. This 'sixth sense' lets our brain know the relative positions in space of different parts of our bodies. Without it, our brains are lost.

Ian Waterman knows how that loss feels. More than 30 years ago he lost this sense almost overnight, when a flu-like virus damaged the required sensory nerves. His muscles worked perfectly, but he could not control them. "I lost ownership of my body," he says. He could no longer stand, or even sit up by himself, and doctors said he would never be able to do so again. Waterman's condition arose from a disease called acute sensory neuropathy and is so rare that only a dozen or so similar cases are known to the medical literature.

Some neuroscientists are taking a cue from Waterman's experiences and starting to investigate whether robotic devices controlled by thought alone could be integrated with an artificial sense of proprioception. If so, they reason, these 'neuroprosthetics' could be made to work in a much more life-like way. What's more, they hope to gain a deeper understanding of how proprioception works, and how they might be able to manipulate it.

Some months after his virus attacked, Waterman, only 19 years old, was lying in bed applying all his mental energy to the fight for control of his body. He tensed his stomach muscles, lifted his head and stared down at the limbs that seemed no longer to belong to him. He willed himself to sit up.

Concentrated effort

Later, he realized that it was the visual feedback that allowed his body to unexpectedly obey the mental instruction. "But the euphoria of the moment made me lose concentration and I nearly fell out of bed," he remembers.

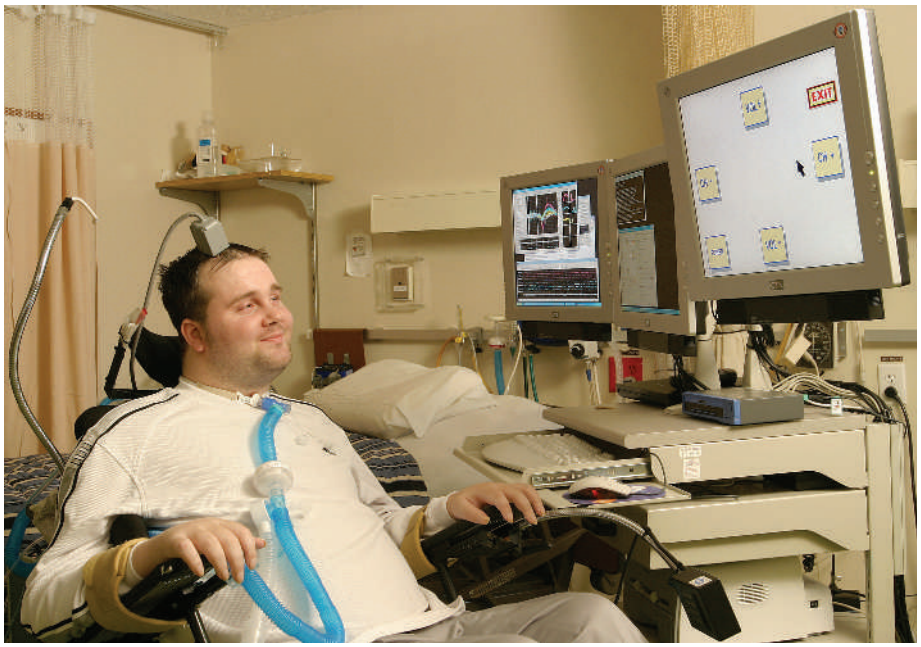
From then on he learnt to compensate for his deficit in proprioception with other forms of sensory feedback to help him understand where his limbs are, and thus control them. It requires constant, intense concentration, but now, despite his profound impairments, he can manage fairly normal movements. Most of the input that he relies on is visual — standing up with his eyes closed is still nearly impossible — but he can also tune in to the tug of a jacket sleeve to work out the direction his arm is moving. Or to the cool air on his armpit when he raises his arm in a loose shirt. Neuroprosthetic engineers are realizing that many sensory feedback signals could be similarly harnessed.

A neuroprosthetic is more accurately called a brain–machine interface. Hundreds of electrodes, fixed into tiny arrays, are placed in or on the surface of the cortex, the thin, folded outer surface of the brain that controls complex functions including the organization of movement. The electrodes record the electrical signals from the cortex's neurons and these are translated by a computer algorithm and used to drive specific actions — the movement of a cursor on a computer screen, for example, or of an artificial limb.

In this issue of *Nature*, two papers[1–3] demonstrate dramatic progress in the area. A team consisting of John Donoghue's group, based at Brown University in Rhode Island and Cybernetics Neurotechnology Systems in Foxborough, Massachusetts, implanted 96 electrodes into Matt Nagle's motor cortex, the brain region that processes information about move-

©2006 Nature Publishing Group

R. FRIEDMAN



Mind control: Matt Nagle's neuroprosthetic lets him move a cursor using thought alone.

ment. Nagle is a quadriplegic patient and the first human volunteer to reach this advanced stage of testing (see picture, above). Hooked up to computers and attended by a team of technicians, Nagle could move a cursor to issue different instructions — for example, to open e-mails or turn down the television.

Krishna Shenoy's group at Stanford University, California, has done similar work in a non-paralysed monkey's premotor cortex, the area of brain where the animal's movement-related 'intentions' are generated. Using a new algorithm, the team's brain–computer interface produced results four times faster and more accurate than previously seen.

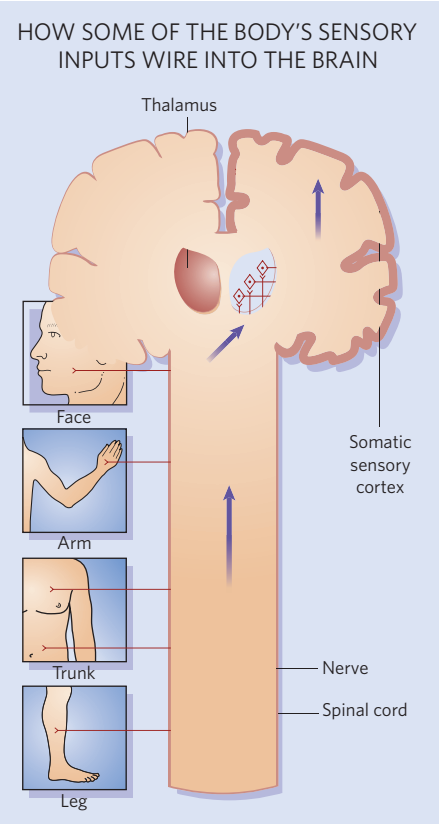
Closing the loop

The two papers show how closely neuroprosthetics are approaching medical reality. But although moving a computer cursor by thought alone may be dazzling, scientists have long-term ambitions to make neuroprosthetics reproduce more complex functions. Could patients direct a robotic arm to pick up a coffee cup, for example? "For this, the devices need to deliver feedback to the brain — we need to close the loop," says Daofen Chen, director of the neural prosthesis programme at the US National Institute of Neurological Disorders and Stroke in Bethesda, Maryland.

The brain's sensory cortex receives signals — proprioception, touch, pain and so on — from the body (see graphic), and in response constantly modifies its movement-related commands. The current generation of output-only neuroprosthetics are open-loop systems — with more limitations than Ian Waterman, who can at least use visual, temperature and tactile feedback. "Brain–machine interfaces will have to become interactive," says Chen. "But now that we would like to exploit it, we realize we know next to nothing about sensory input."

A handful of researchers is starting to try to work out where and how to stimulate the sensory nervous system to reproduce the sorts of information that a limb might send to the sensory cortex. It is early days: none of their work is published. And as so little is known about the system, there is no obvious place to start.

Theoretically, the 'where' could be the nerves running from the limb into the spinal cord, or the spinal cord itself (see graphic). Or it could be higher — in the brain's thalamus, where incoming sensory signals are integrated and redirected to the appropriate part of the cortex, or the sensory cortex itself.



The 'how' refers to the design of the electrical signals to be fed into the cortex. These could mimic the sensory system's natural nerve impulses, based on parameters such as frequency and amplitude. Or they could involve creating artificial signals that the sensory cortex is able to distinguish, in the hope that the brain can be trained to associate particular signals with particular parameters.

Once scientists have worked out how best to encode the signals, the idea would be to place sensors on artificial limbs to generate signals representing proprioceptive information such as angle of joint, vibration, force of grip — and other sensory information that Waterman has found helpful, such as temperature.

Trained brain

Most in the field have a hunch that the signal will not have to mimic neural activity perfectly. The brain can, after all, cope with the very unphysiological signals generated by the most successful brain–machine interface to date: the cochlear transplant. Already, some 110,000 profoundly deaf people have been implanted with the device, according to the US National Institutes of Health. The implant sits in the inner ear and interfaces with the auditory nerve. Its signals are totally artificial, and, at first, recipients can make nothing of the noise. But the auditory cortex, it turns out, is highly adaptable. With appropriate training, it can quickly learn to associate particular codes with particular sounds, so that transplant recipients can learn to follow conversations with ease.

"When the concept of stimulating the auditory nerve emerged in the 1970s, people said it would be impossible to generate the right electrical signal to the brain," says Shenoy. "But it turned out that you don't have to get it perfect, just close enough for the brain to do its own fine-tuning." On the other hand, one does not want to burden patients with having to learn too much, says John Chapin a physiologist at the State University of New York Health Science Center in Brooklyn, and a pioneer in using neural activity to control robots. "Ideally we should aim to mimic the natural signal as closely as possible," he says (see 'Voyagers in the cortex').

For now, whatever works will be good. "We don't know if it will turn out to be possible to incorporate sensory information but we are going to try," says neuroscientist Andrew Schwartz of the University of Pittsburgh, an expert in brain–computer interface technology for the control of robotic arm movement.

Schwartz is working with Douglas Weber, a bioengineer at Pittsburgh who is developing a model for studying sensory input. This involves using electrodes to stimulate the sensory nerves from the limbs of an anaesthetized cat at the point just before they enter the spinal cord, and simultaneously recording

ADAPTED FROM: M. J. T. FITZGERALD & J. FOLAN-CURRAN *CLINICAL NEUROANATOMY AND RELATED NEUROSCIENCE*

©2006 Nature Publishing Group

Voyagers in the cortex

If paralysed people can now use electronic implants to move a computer cursor by thought, could similar implants make the deaf hear, or the blind see? Not any time soon, say researchers. The brain's visual and auditory cortices, which process vision and hearing, respectively, are extremely complex — and early experience has been discouraging.

A few decades ago neurologists tried stimulating combinations of neurons, in the hope that patients would perceive complex shapes or sounds. The patients perceived strong signals, but they were unstructured — flashing light spots, or hissing sounds[4,5].

Frank Ohl of the Leibniz Institute for Neurobiology in Magdeburg, Germany, believes that earlier studies may have failed because the cortex is too complicated to be stimulated using the same simple signalling that works on the peripheral nerves that feed into it. "It is constantly active, and 96% of that activity is internal — different parts of the cortex exchanging information with each other," he explains.

Ohl's team has been monitoring this internal chatter while experimenting with different patterns of electrical stimulation. He hopes to find a way of delivering just the right kinds of signals at just the right point amid the chatter when the cortex is able to 'listen' to them. Fiendishly complicated, but Ohl has already shown that his chosen subjects, gerbils, can discriminate between rising and falling tones whether they are presented as sound, or as direct stimulations to the cortex that create the same pattern of cortical activity as the sound produces[6]. He also has unpublished evidence that the gerbils become more accurate in distinguishing the signals when he takes the chatter into account.

A few weeks ago, Ohl's university signed an agreement with neurosurgeon Volker Sturm from the University of Cologne to move the project on to humans. Epileptics sometimes have electrodes implanted in the areas of their brain believed to generate their seizures. If these areas happen to be in or close to the auditory cortex, Ohl will attend the operations and repeat some of the animal experiments on patients willing to take part.

Ed Tehovnik, a neuroscientist at the Massachusetts Institute of Technology, is taking a similar approach in the visual cortex in monkeys. He thinks earlier attempts may have failed because they did not consider how important the cortex's detailed architecture is to function. Where and how you interface with this part of the brain is critical, he says.

This may all be a long way from restoring sight or hearing, but it will at least establish the physiological principles on which sensory neuroprostheses may be based. "In the meantime we will find out a lot about how the brain works," says Ohl. **A.A.**

from neurons in the sensory cortex. Weber will then repeat the recording, only this time manually moving the cat's limbs, instead of stimulating their nerves with electrodes. He will then compare the pattern of neural activity in the cortex in the two situations and see whether he can mimic the patterns he sees in response to passive movements with artificial stimulation.

"Not everyone agrees, but my gut feeling is that we will be more successful if we stimulate outside of the central nervous system," says Weber. "At more central points there will be greater convergence of different inputs and I guess it would be hard to get clean signals".

Lee Miller, a neurophysiologist from Northwestern University in Chicago, agrees that Weber could be right, but is nevertheless approaching the problem from the top. Although useful for study, stimulating nerves from peripheral areas of the body such as limbs will not work for a patient whose spinal cord is severed.

Working in monkeys, Miller's group is electrically stimulating the part of the cortex that processes proprioception, and recording neuronal activity in the motor cortex at the same time. Miller hopes this will eventually let him design stimulation patterns that can imitate the brain's own processing of proprioceptive signals, much in the way that Weber is designing signals to imitate processing of movement. Monkeys will be trained to move a 'virtual' arm, created on a computer screen by computer algorithms fed by both recordings from the motor cortex and the simulated proprioceptive feedback.

It's a complex experiment, he admits, which will probably take up to five years to get working optimally. But he draws hope from his group's finding, presented at the 2005 Society of Neuroscience meeting in Washington DC, that the monkeys can recognize and distinguish between high- and low-frequency stimulation.

Chapin's set-up is equally ambitious. He also works on monkeys but his chosen target is the thalamus, the brain's junction box for sensory input. "The higher you go in the brain, the more complex and abstract things become," he says. "It is hard to know if you are stimulating something precise." In his experimental system, Chapin electrically stimulates the area of the thalamus that relays touch-related signals. Simultaneously he records in the areas of the sensory cortex that process tactile information. The monkeys meanwhile, have one arm strapped down and one free. They have been taught to point with their free hand to an area on their immobilized arm that they 'feel' is being touched. "We have found that we can produce a sort of 'natural response' in the cortex when we stimulate in the thalamus," he says. The response matches that produced normally when a specific part of the monkey's arm is touched. Chapin plans to extend his investigations to study proprioception in the same way.

The papers on brain-machine interfaces by Donoghue and Shenoy[1,2] seem like science fiction becoming reality. The next step — trying to introduce sensory input into brain–machine interfaces — may appear at first glance to be as fanciful as the *Six Million Dollar Man*. But few neuroscientists could seriously doubt their theoretical potential. As experience with the first generation of neuroprosthetics shows — it's a question of understanding how the system works. ■

Alison Abbott is *Nature*'s senior European correspondent.

CONDÉ NAST ARCHIVE/CORBIS

Daily operation: the light traces of a person getting dressed show that even simple tasks require complicated movements.

1. Hochberg, L. R. *et al. Nature* **442,** 164–171 (2006).
2. Santhanam, G., Ryu, S. I., Yu, B. M., Afshar, A. & Shenoy, K. V. *Nature* **442,** 195–198 (2006).
3. Scott, S. H. *Nature* **442,** 141–142 (2006).
4. Brindley, G. S. & Lewin, W. S. *J. Physiol. (Lond.)* **196,** 479–493 (1968).
5. Dobelle, W. H., Stensaas, S. S., Mladejovsky, M. G. & Smith, J. B. *Rhinol. Laryngol.* **82,** 445–463 (1973).
6. Ohl, F. W. *et al. Nature* **412,** 733–736 (2001).

©2006 Nature Publishing Group

NATURE|Vol 442|13 July 2006

The start of the world as we know it

Plate tectonics has created oceans and pushed up mountain ranges. But when did the process that shapes the planet get going? **Alexandra Witze** joins the geologists debating the issue.

On a sunny, breezy day in the Wind River Mountains of Wyoming, a group of geologists are peering intently at a dark ridge of rock. Some 2.7 billion years ago, these rocks were alive with volcanic fire. Today, they jut out of a mountainside like the spiny tail of a sleeping dragon.

This rock, says Kevin Chamberlain, a geologist from the University of Wyoming in Laramie, could be a special kind — a lava called a komatiite. Today, Earth's interior is too cool to produce this particular rock; 2.7 billion years ago, the hot lava would have run like water over the barren landscapes.

But as the other geologists chip off fresh layers and scrutinize them through hand lenses, murmurs of dissent start to grow. Most geologists have never seen a komatiite; they are found almost exclusively among rocks of the Archaean era, which are more than 2.5 billion years old and thus very rare. But these men and women are experts in the truly old. And on the Wyoming hillside few of them are convinced that they are seeing the rock textures typically found in komatiites.

The scene brings home the difficulties of trying to study the early Earth — there aren't many old rocks to look at, and those that are around are often so altered, chemically and physically, as to be nearly indecipherable. But as if to recompense those who study them, such ancient rocks, particularly of Archaean age, offer geologists great rewards. It is in the Archaean that the first earthly ecosystems are found, with their clues to life's earliest days on the planet. And it is in the Archaean that scientists can look for the beginnings of plate tectonics.

Plate tectonics is the grand unified theory of geology. Everything we see today, from the abyssal plains of the oceans to the heights of the Himalayas, is shaped by plate tectonics. As far back as there has been complex life — and perhaps even before — continents have come together and moved apart in a dance that has altered climates and geographies, opening up new possibilities for life and sometimes closing down old ones.

But it may not always have been so. Plate tectonics is driven by Earth's heat and constrained by the physical and chemical properties of the crust and mantle. The further back in time you go, the more different these things are likely to have been. It's been argued that on the early Earth, crustal plates, floating on a heat-softened layer of material beneath, would have simply been too thick and buoyant to get dragged beneath each other as they are today. And the greater temperature of the early Earth's innards would probably have made them move in very different patterns from those typical of today's tectonics.

A. WITZE

Ancient rocks, such as those in the Wind River Mountains, could help determine when plate tectonics began.

On other Earth-like planets there's no evidence for today's plate tectonics. Planets do not have to work this way, and there was probably a time when this one didn't. "You don't just make a silicate planet and plate tectonics starts," says Robert Stern, a geologist at the University of Texas, Dallas. "Something special has to happen."

Dynamic planet

The nature of that special something cuts to the discipline's philosophical heart. Since the early nineteenth century, geology has been ruled by the principle of uniformitarianism — that the planet operates on unchanging laws, and that the present can be used as a key to the past. But how can that approach hold up when a science from a world where plate tectonics explains more or less everything is applied to a world that may have lacked it? How can you understand ancient rocks when you do not know what processes put them there?

The geologists clustered around the possible komatiites in the Wyoming hills had gathered to discuss these questions. Their visit to the mountain ridge had been organized as part of a June conference held in Lander, Wyoming. The specific aim of the meeting was to try to fix a date for the onset of plate tectonics: the earliest possibility is pretty much straight after the planet formed, about 4.5 billion years ago; the latest is just 1 billion years ago.

To help them decide, the scientists brought to the table data from an array of disciplines. Geochemistry can help clarify the temperature and pressure at which Archaean rocks formed. Fragments of zircon crystals dated even earlier

©2006 Nature Publishing Group

— to the Hadean eon, which stretches from about 3.8 billion years ago to the planet's birth — can provide hints about what Earth's surface environment was like back then. Palaeomagnetic studies can show how land masses moved across latitudes. And structural geology can identify features that, in today's world at least, seem to be indicative of plate tectonics. But in all these approaches, as with the komatiites, age makes the picture hard to discern.

Crashing continents

Scant and difficult-to-interpret evidence presents one set of problems; slippery definitions present another. Plate tectonics has lots of constituent parts. It's not just a theory of how things move, but of how they are made and from what. For example, explanations for different sorts of volcanism in different settings also explain why the mineral make-up of continental crust and the crust beneath the oceans is so different.

Working out which attributes are essential to the theory, and which incidental, is not easy. The 65 attendees at the Wyoming conference came up with 18 different definitions of plate tectonics. Three components, most agreed, were key: there must be rigid plates at the surface of the Earth; those plates must move apart through ocean spreading, with new crust being made where the sea floor pulls apart; and the plates must on occasion dive beneath each other at subduction zones (see graphic).

The problem is that Earth could display one or even two of these properties without necessarily having a system like that described by modern plate tectonics (see 'A world without tectonics'). Take rigid plates. Palaeomagnetic and other studies show that sections of Earth's crust moved relative to each other in the Archaean, just as modern crustal plates do. But ice floes on a polar sea move in the same way, points out geophysicist Don Anderson of the California Institute of Technology in Pasadena — and those ice floes aren't experiencing plate tectonics.

Of the three, it seems subduction is closest to being diagnostic of plate tectonics. Subduction is the process by which one crustal plate slips beneath another, to be recycled into the mantle. Subduction requires rigid plates, and as it involves the destruction of crust, new crust must be created elsewhere, presumably at oceanic spreading ridges (see graphic); otherwise, continental crust would eventually disappear. Some argue that this means plate tectonics should date further back than the earliest firm evidence for subduction.

A dramatic use of this argument is that made by Stern. In a paper published last year, he took an extreme position, proposing that Earth has been free of plate tectonics for almost four-fifths of its life, with the system we see today starting up only a billion years ago[1]. He had two lines of argument. One was that plate tectonics could not begin until Earth's crust was cool enough, and that barrier was only passed about a billion years ago. The other was that the only reliable evidence for subduction on the early planet came from a period more recent than that.

> **"You don't just make a silicate planet and plate tectonics starts. Something special has to happen."**
> — Robert Stern

Stern points to the geological record of three types of rock. Ophiolites are distinctive sections of the ocean crust that gets mashed up, often through subduction, on the edges of continents. Stern argues that very few of these rocks are more than a billion years old. Metamorphic rocks called blueschists, produced by squeezing the basalt from which oceanic crust is made at high pressures but not very high temperatures, are being made in today's subduction zones; none, Stern says, has been found that is more than 800 million years old. And rocks from 'ultra-high-pressure terranes' of the sort produced where one plate rides over another are at most 630 million years old.

He also makes a more general point. A dramatic shift, such as the introduction of plate tectonics, must have had huge planetary consequences. And between 780 million and 580 million years ago, Stern says, there was a series of glaciations, some very extreme — giving rise to the term 'snowball Earth'. "It was a wild time of change," says Stern. "The biosphere was out of control." On the basis that dramatic effects require dramatic causes, he argues that the introduction of plate tectonics, and with it an increase in planet-cooling volcanic eruptions, might have precipitated the great glaciations.

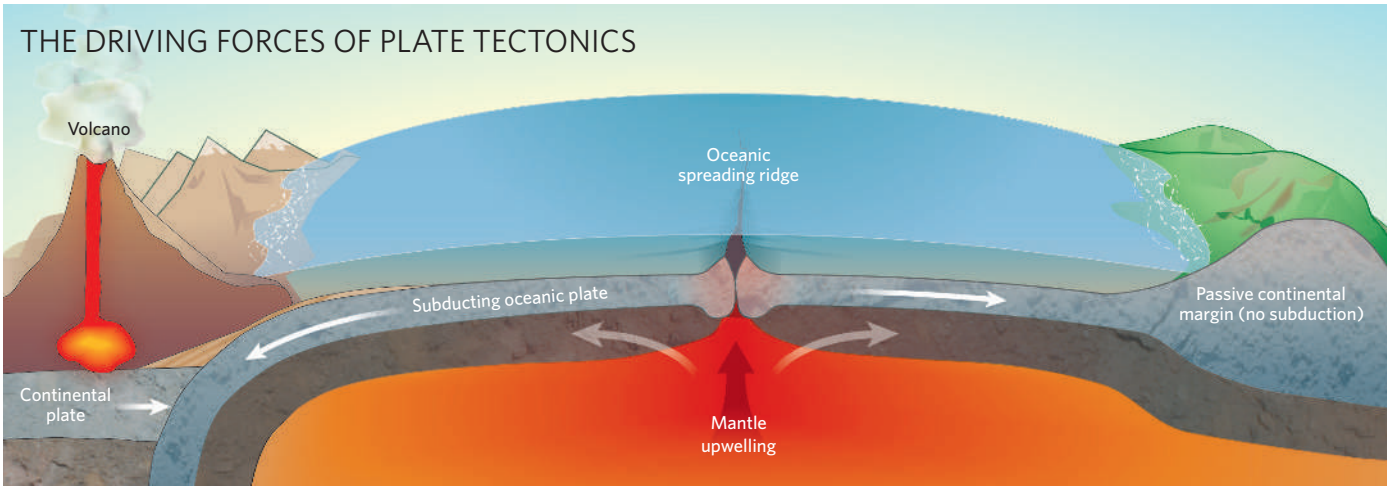
An age gone by

After reading Stern's arguments, Alfred Kröner of the University of Mainz in Germany fired off a rebuttal. He argues that there's plenty of evidence for plate tectonics stretching back at least 3.1 billion years[2] — including geochemical work, seismic images of the 'sutures' where colliding continents join and, indeed, a few ancient ophiolites. "I believe we can see these features all the way back" — possibly all through the Archaean, says Kröner.

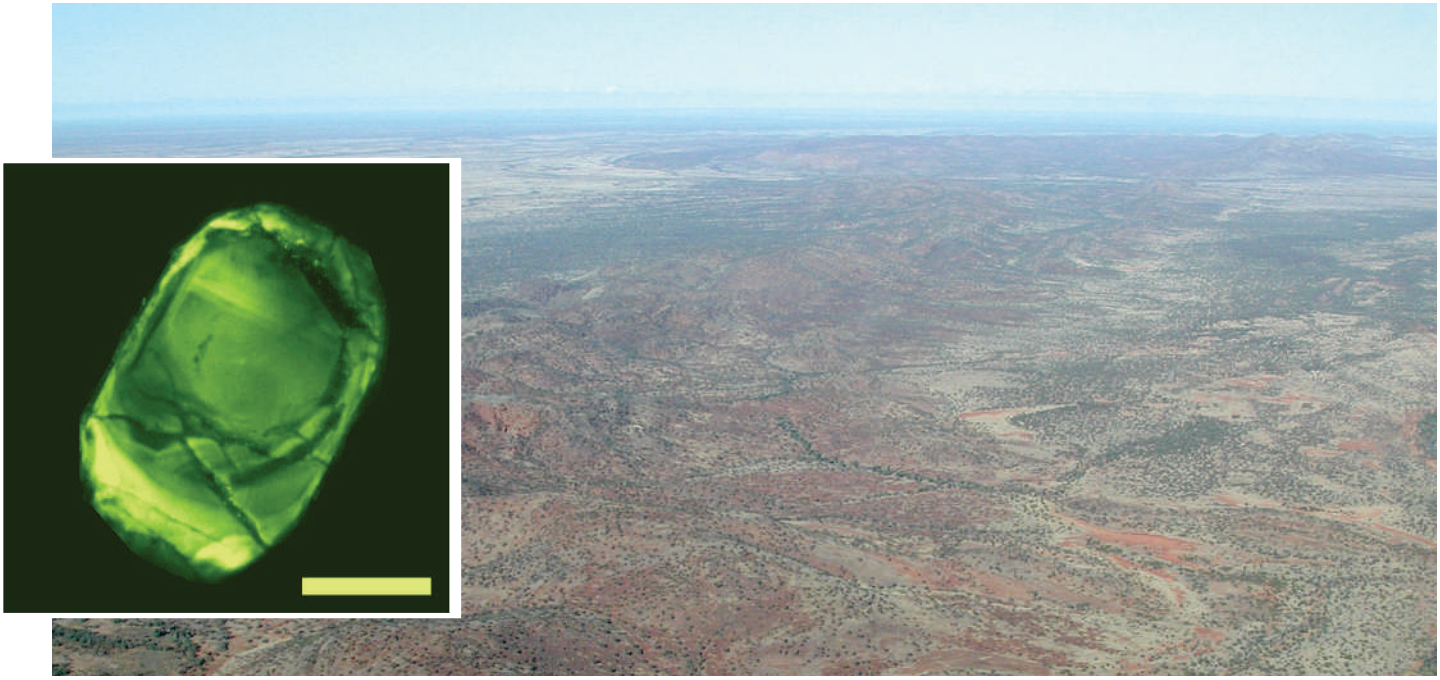
The exchange of papers led to the Wyoming conference. "It was overdue," says Kröner. "Nobody ever talks to one another." In Wyoming, they did: palaeomagnetists clustered around a white board with field geologists; geophysicists sat down for a beer with geochemists.

Some of the newly shared data favour an early start for plate tectonics. Geoff Davies, a modeller at Australian National University in Canberra, presented work suggesting that one of the biggest stumbling blocks to an early start may have been removed. In the early 1990s, computer models created by Davies and others suggested that the crust on the early Earth would have been too thick and buoyant to get dragged down beneath another plate during subduction. But new simulations, using more sophisticated calculations, suggest that the crust may have been thinner than once thought[3] — as thin as 4 kilometres or less — which would be thinner than today's crust. "Maybe plate tectonics on the early Earth was viable after all," says Davies.

In other cases, recent findings overturned evidence for early plate tectonics. In 2001, a team reported that an ophiolite from Dongwanzi, China, was 2.5 billion years old — mak-



©2006 Nature Publishing Group



M. HARRISON

Clues to the past: zircon crystals (inset) in the Jack Hills of Western Australia have been used to argue for an early start to plate tectonics.

ing it by far the oldest such subduction remnant yet discovered[4]. Now Guochun Zhao, of the University of Hong Kong, has re-dated those rocks, giving them an age of just 300 million years.

Timothy Kusky of St Louis University in Missouri, who led the original study, says that Zhao took samples from a part of the rock already known to be much younger than the main part of the ophiolite. But several attendees at the meeting said they found Zhao's data convincing. If true, it would pull the earliest evidence for ophiolites at least half a billion years towards the present, leaving the Archaean an ophiolite-free zone.

The Chinese ophiolite isn't the only evidence that is getting fresh scrutiny. For a while two independent groups have been quietly warring over the significance of a pile of ancient zircons from the Jack Hills region of Western Australia. The zircons are crystals that formed in the Hadean and later became incorporated into younger rocks.

Last year in *Science*[5], geochemist Mark Harrison of the University of California, Los Angeles, and colleagues used the Jack Hills zircons to argue that continental crust was present 4.4 billion to 4.5 billion years ago. The evidence comes in the form of hafnium isotope ratios in the zircon crystals, which preserve signals of the lighter minerals typical of continental crust. The data also suggest, Harrison argues, that that crust was being recycled down into the mantle by 4.4 billion years ago — perhaps though a process similar to plate tectonics.

Simon Wilde of the Curtin University of Technology in Perth, Australia, isn't so sure. "You have to be very careful with these rocks," he says. Measuring one spot on a crystal, as opposed to another, can yield very different hafnium values that lead to very different interpretations, he says. Wilde argues that the zircons should be interpreted more conservatively — that the evidence points to there being some continental crust, but not plate tectonics and its associated recycling, by 4.4 billion years ago[6].

Ground forces

Such differences of interpretation make the problem of solving when plate tectonics began extremely difficult. In many cases, data can be interpreted in several completely different ways — all of which may seem valid.

For instance, another Australian geologist presented seemingly convincing evidence that plate tectonics had begun by 3.3 billion years ago in Western Australia, based on the very different histories of two sections of an ancient rock formation called the Pilbara. Hugh Smithies of the Geological Survey of Western Australia says that the eastern part of the Pilbara, between 3.5 billion and 3.2 billion years old, "shows no clear evidence for modern-style plate tectonics". It contains some geochemical markers that suggest subduction, but they could just as easily be explained by hot upwellings of rock known as mantle plumes or other non-tectonic phenomena.

In contrast, looking at the western part of the Pilbara — which is 3.3 billion to 3.0 billion years old — Smithies sees plenty of evidence for plate tectonics. There are geochemical signatures that cannot be explained by other factors, and the

A. J. VAN DER VELDEN

Time to talk: Earth scientists gathered at a meeting in Wyoming to present diverse data on the early Earth.

©2006 Nature Publishing Group

A. POLAT

Slow work: geologists hunt for structural signatures in rocks that can only be explained by plate tectonics, to try to identify when the process started.

rocks show features that hint that plates had interacted along their edges. Smithies thinks the western Pilbara contains the remains of an oceanic arc — the sort of line of islands, such as the Aleutians of Alaska, that are characteristic of some oceanic subduction zones[7].

But then along came Julian Pearce of Cardiff University in Wales, who argued that each of the geochemical markers in the western Pilbara can be explained by other phenomena, such as magmas with an unusual amount of water in them, or crustal material from different places getting mixed up. The various researchers are hoping to settle the matter with a field trip. An excursion is already planned for next year, to re-examine the evidence for plate tectonics in the western Pilbara.

Field trips don't always resolve things. In the Wind River Mountains, the meeting attendees continued to argue about plate tectonics as they hiked from outcrop to outcrop. But a week of communing at the conference and under the high mountain sun brought them toward a consensus of sorts.

Meeting organizers polled the attendees twice on when they thought plate tectonics began. At the beginning of the meeting, guesses were spread over more than 3 billion years of Earth history. At the end, a closing ballot showed that many had begun to push their thinking further back into the past; a majority of attendees voted for plate tectonics having started between 3 billion and 4 billion years ago.

Kent Condie, one of the meeting organizers, calls that a success. "We've got a majority favouring a definition and approach," says Condie, of the New Mexico Institute of Mining and Technology in Socorro, New Mexico. "Sure, there will be a minority point of view."

At the conference, that minority pretty much constituted Stern. By the end of the meeting, he remained the one person voting for a start to plate tectonics at 1 billion years ago. "It's not a simple question," he maintains. And on that, at least, others agree.

Michael Brown, a geologist at the university of Maryland in College Park, doesn't endorse Stern's late start. But he does think that the nature of plate tectonics changed around that time. In a paper in press in *Geology*[8], Brown suggests that there have been two styles of plate tectonics: the modern kind that we see today, and an earlier version that lasted from about 2.7 billion to 700 million years ago. Evidence for the earlier style, he says, comes from minerals that are typical of higher-temperature, lower-pressure environments; these suggest a hotter Earth where plates did not subduct beneath each other to great depths and pressures. Minerals characteristic of high-pressure environments typify the later style. The properties of these minerals suggest to him that true plate tectonics, in which one plate subducts deeply beneath another, did not begin until about 700 million years ago.

And there is a possible further complication. Geophysicist Paul Silver, of the Carnegie Institution of Washington, raised the notion that plate tectonics may have started and stopped several times during Earth's history. This is also an idea that Stern is comfortable with — he uses it to explain the presence of a small number of ophiolites about 2 billion years ago.

An 'intermittent approach' would be a wonderful way to reconcile things — but it takes geology even further from the comforting realm of uniformitarianism, into a world where the most basic principles come and go in fits and starts. ■

Alexandra Witze is *Nature*'s chief of correspondents for America.

A WORLD WITHOUT TECTONICS

With so much uncertainty over Earth's early history, not many geologists are willing to imagine entirely different ways in which the world might have worked before plate tectonics. One of the brave few is Jean Bédard, a geologist at the Geological Survey of Canada.

At a June conference held in Lander, Wyoming, Bédard presented one of the few fully worked out accounts of how a pre-plate-tectonic Earth might have worked. In his model, hot upwellings of rock known as mantle plumes partly melt the crust above them. This melting distills the crust, producing material from which light, continental-style crust is made.

The material left behind as the melt is creamed off — denser than it was before the distillation — then detaches itself from the crust and sinks back into the mantle. There, it would mix with more mantle material — and the whole process would start all over again[9].

Bédard himself admits that he has no idea if his proposal is right. But it is, he says, a detailed alternative theory to plate tectonics, and one that can be tested with further studies. **A.W.**

1. Stern, R. J. *Geology* **33,** 557–560 (2005).
2. Cawood, P. A., Kröner A., Pisarevsky, S. *GSA Today* **16,** 4–11 (2006).
3. Davies, G. F. *Earth Planet. Sci. Lett.* **243,** 376–382 (2006).
4. Kusky, T. M., Li, J-H. & Tucker, R. D. *Science* **292,** 1142–1145 (2001).
5. Harrison, T. M. *et al. Science* **310,** 1947–1950 (2005).
6. Valley, J. W. *et al Science* **312,** 1139a (2006).
7. Smithies, R. H. *et al. Gondwana Res.* (in the press).
8. Brown, M. *Geology* (in the press).
9. Bédard, J. *Geochim. Cosmochim. Acta* **70,** 1188–1214 (2006).

©2006 Nature Publishing Group

Misconduct: forum should not be used to settle scores

SIR — Although China is developing its science and technology at an unprecedented speed, scientific misconduct is a serious issue, as you have highlighted in your Special Report "Named and shamed" (*Nature* **441,** 392–393; 2006).

Shi-Min Fang, the webmaster of New Threads (www.xys.org), has defended, in Correspondence (*Nature* **441,** 932; 2006), this website's role in disclosing scientific misconduct on occasions when the authorities have ignored whistleblowers.

Like many other Chinese scientists working overseas, I care very much about scientific misconduct in China. However, I have also been concerned for a long time about the quality of articles published on New Threads. Often, I find that there are few facts and little investigation behind the accusations, and that many articles are mixed with assumptions and personal attacks on named scientific researchers.

One such example is that of Hualiang Jiang, a principal investigator working at the Shanghai Institute of Materia Medica. Because I work in a similar field, I am familiar with Jiang's work and publications, although I have never met him. New Threads contains several articles (urls provided) attacking Jiang personally, using many insulting words such as "idiot". It seems that some of the articles were written by someone who may have been an unsuccessful job candidate at Jiang's institute.

Disclosing scientific misconduct is not simply about free speech, as claimed by Fang. It is also about being professional, objective and serious. Only verified facts should be published on the website, if it is claiming to monitor incidents of scientific misconduct. It should not be used for unsubstantiated attacks in the name of free speech, not only because of the personal and professional effects on the scientists concerned, but also because readers, especially young students, could be misled.

Guosheng Wu
700 Lower State Road, North Wales, Pennsylvania 19454, USA

Misconduct: China needs university ethics courses

SIR — Your Special Report "Named and shamed" (*Nature* **441,** 392–393; 2006) and Editorial "Finding fraud in China" (*Nature* **441,** 549–550; 2006) report that scientific misconduct has become rampant in China, especially in universities. I would like to add my view to those of previous correspondents (*Nature* **441,** 932; 2006).

Misconduct is hampering the sound development of science in the nation's higher-education system. If scientific misconduct cases are not handled by the university (or other concerned authority), and much-needed outside supervision is not available, then each occasion that comes to light damages the academic reputation of the university concerned, the whistleblower and the person accused.

Unfortunately, serious and justified investigations of suspected fraud have been largely ignored by China's universities, with the exceptions of the prestigious Tsinghua University and Shanghai Jiaotong University, each of which has recently dismissed a professor (returned from abroad in each case) for fabricating research achievements or results. However, details of these investigations have not been disclosed, so other universities cannot learn from them.

Obviously, it is the university involved in a fraud case — not the Ministry of Education, the media, websites, journals or newspapers — that has the power to dismiss or demote the accused, if guilty. A mechanism is needed to deal with such eventualities.

To cope with embarrassing situations such as those currently being highlighted in the media, I suggest that editorials and articles on the subject in science journals such as *Nature* and *Science* should be used as materials for teaching a course of research ethics to students in China's universities. Access to case studies being taught in scientific ethics courses elsewhere would also be valuable. Our universities should play a key part in fighting scientific misconduct, and every honest Chinese professor should make a contribution to such courses as part of providing a complete university education.

Qizhi Wang
Department of Mechanics, College of Architecture and Environment, Sichuan University, Chengdu, Sichuan 610065, China

Rushed decision on collider would limit useful options

SIR — Your Editorial about the International Linear Collider (ILC), "Making collider endorsement count" (*Nature* **440,** 1089; 2006), states that CERN has insisted "that decisions about the siting of the ILC be delayed until an accelerator technology it is trying to develop is ready". This is untrue.

CERN's position is that no irreversible decision on building a linear collider should be taken until the end of the decade. By that time, the Large Hadron Collider — the particle-physics community's current flagship facility — will have produced its first results, a full technical design for the ILC will be ready, and we will know whether the technology for a more powerful compact linear collider (CLIC) is feasible. These ingredients will allow the global particle-physics community to take an informed and responsible decision on its future. Moreover, funding is simply not available on a shorter timescale for such a huge project to begin.

You report that some in the community find CERN's position "self-serving". Yet this is by definition not possible, given that it implements the decisions taken by its 20 member states. The organization's position on the ILC is necessary to ensure the optimum use of public resources in this exciting area of research. Contrary to the implication in your Editorial, CERN has accelerated research and development on CLIC precisely so as not to delay the decision on the ILC.

Also contrary to the implication in your Editorial, CERN contributes to the ILC by providing personnel and material resources, and by participating in European networks for ILC research and development.

Robert Aymar
CERN, Geneva 23, CH-1211, Switzerland

Speaking for Taiwan about colours, maps and politics

SIR — In your News Feature "Forward planning" (*Nature* **440,** 987–989; 2006), you published a map of East Asia in which Taiwan and China are both coloured yellow, indicating that they comprise one country. Yet Taiwan and China have been governed separately since 1949, when the Chinese communists ousted the nationalists from the mainland in a civil war.

Since then, Taiwan (officially called the Republic of China) has undergone a complete process of democratization, while China (officially called the People's Republic of China) has continued under single-party communist rule. This is a reality that cannot be ignored. If a reporter from *Nature* needed to visit Taiwan on assignment, for example, he or she could not apply for a visa at the Chinese embassy.

I hope that in future *Nature* will take into account the fact that Taiwan and China are governed separately as two distinct countries.

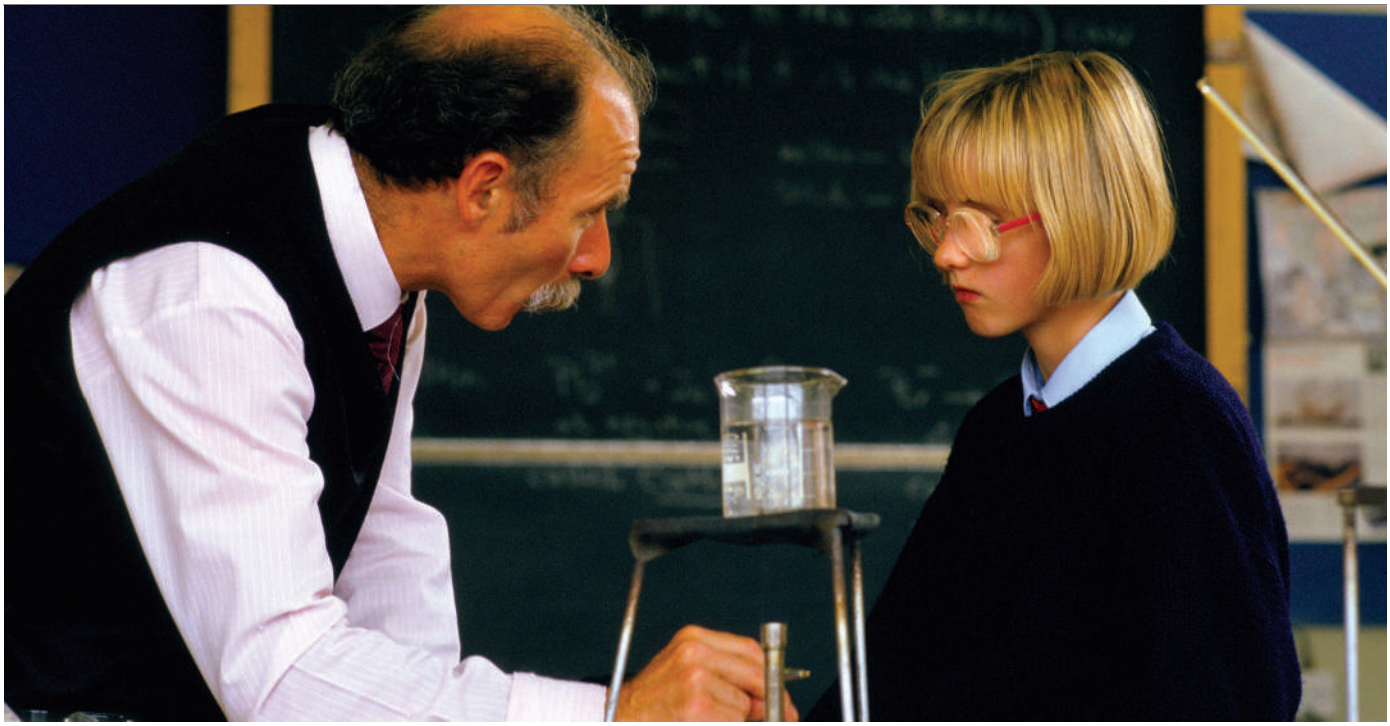
Michael Chen
Taipei Representative Office in the United Kingdom, 50 Grosvenor Gardens, London SW1W 0EB, UK

***Nature*'s long-standing policy is to abide by the United Nations' position on the status of Taiwan — Editor, *Nature*.**

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.

©2006 Nature Publishing Group

COMMENTARY



M. GOLDWATER/ALAMY

Does gender matter?

The suggestion that women are not advancing in science because of innate inability is being taken seriously by some high-profile academics. **Ben A. Barres** explains what is wrong with the hypothesis.

When I was 14 years old, I had an unusually talented maths teacher. One day after school, I excitedly pointed him out to my mother. To my amazement, she looked at him with shock and said with disgust: "You never told me that he was black". I looked over at my teacher and, for the first time, realized that he was an African-American. I had somehow never noticed his skin colour before, only his spectacular teaching ability. I would like to think that my parents' sincere efforts to teach me prejudice were unsuccessful. I don't know why this lesson takes for some and not for others. But now that I am 51, as a female-to-male transgendered person, I still wonder about it, particularly when I hear male gym teachers telling young boys "not to be like girls" in that same derogatory tone.

Hypothesis testing

Last year, Harvard University president Larry Summers suggested that differences in innate aptitude rather than discrimination were more likely to be to blame for the failure of women to advance in scientific careers[1]. Harvard professor Steven Pinker then put forth a similar argument in an online debate[2], and an almost identical view was elaborated in a 2006 essay by Peter Lawrence entitled 'Men, Women and Ghosts in Science'[3]. Whereas Summers prefaced his statements by saying he was trying to be provocative, Lawrence did not. Whereas Summers talked about "different availability of aptitude at the high end," Lawrence talked about average aptitudes differing. Lawrence argued that, even in a utopian world free of bias, women would still be under-represented in science because they are innately different from men.

> "Few tragedies can be more extensive than the stunting of life, few injustices deeper than the denial of an opportunity to strive or even to hope, by a limit imposed from without, but falsely identified as lying within."
> — Stephen Jay Gould

Lawrence draws from the work of Simon Baron-Cohen[4] in arguing that males are 'on average' biologically predisposed to systematize, to analyse and to be more forgetful of others, whereas females are 'on average' innately designed to empathize, to communicate and to care for others. He further argues that men are innately better equipped to aggressively compete in the 'vicious struggle to survive' in science. Similarly, Harvard professor Harvey Mansfield states in his new book, *Manliness*[5], that women don't like to compete, are risk adverse, less abstract and too emotional.

I will refer to this view — that women are not advancing because of innate inability rather than because of bias or other factors — as the Larry Summers Hypothesis. It is a view that seems to have resonated widely with male, but not female, scientists. Here, I will argue that available scientific data do not provide credible support for the hypothesis but instead support an alternative one: that women are not advancing because of discrimination. You might call this the 'Stephen Jay Gould Hypothesis' (see left). I have no desire to make men into villains (as Henry Kissinger once said, "Nobody will ever win the battle of the sexes; there's just too much fraternizing with the enemy"). As to who the practitioners of this bias are, I will be pointing my finger at women

©2006 Nature Publishing Group

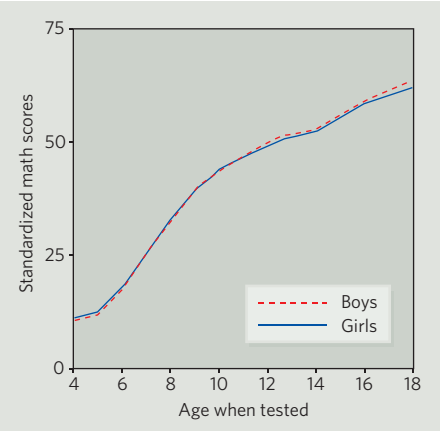


Figure 1 | Maths-test scores for ages 4 to 18. In the United States there is little to distinguish the maths-test scores of boys and girls throughout school.

as much as men. I am certain that all the proponents of the Larry Summers Hypothesis are well-meaning and fair-minded people, who agree that treatment of individuals should be based on merit rather than on race, gender or religion stereotypes.

The sums don't add up

Like many women and minorities, however, I am suspicious when those who are at an advantage proclaim that a disadvantaged group of people is innately less able. Historically, claims that disadvantaged groups are innately inferior have been based on junk science and intolerance[6]. Despite powerful social factors that discourage women from studying maths and science from a very young age[7], there is little evidence that gender differences in maths abilities exist, are innate or are even relevant to the lack of advancement of women in science[8]. A study of nearly 20,000 maths scores of children aged 4 to 18, for instance, found little difference between the genders (Fig. 1)[9], and, despite all the social forces that hold women back from an early age, one-third of the winners of the elite Putnam Math Competition last year were women. Moreover, differences in maths-test results are not correlated with the gender divide between those who choose to leave science[10]. I will explain why I believe that the Larry Summers Hypothesis amounts to nothing more than blaming the victim, why it is so harmful to women, and what can and should be done to help women advance in science.

If innate intellectual abilities are not to blame for women's slow advance in science careers, then what is? The foremost factor, I believe, is the societal assumption that women are innately less able than men. Many studies, summarized in Virginia Valian's excellent book *Why So Slow?*[11], have demonstrated a substantial degree of bias against women — more than is sufficient to block women's advancement in many professions. Here are a few examples of bias from my own life as a young woman. As an undergrad at the Massachusetts Institute of Technology (MIT), I was the only person in a large class of nearly all men to solve a hard maths problem, only to be told by the professor that my boyfriend must have solved it for me. I was not given any credit. I am still disappointed about the prestigious fellowship competition I later lost to a male contemporary when I was a PhD student, even though the Harvard dean who had read both applications assured me that my application was much stronger (I had published six high-impact papers whereas my male competitor had published only one). Shortly after I changed sex, a faculty member was heard to say "Ben Barres gave a great seminar today, but then his work is much better than his sister's."

Anecdotes, however, are not data, which is why gender-blinding studies are so important[11]. These studies reveal that in many selection processes, the bar is unconsciously raised so high for women and minority candidates that few emerge as winners. For instance, one study found that women applying for a research grant needed to be 2.5 times more productive than men in order to be considered equally competent (Fig. 2)[12]. Even for women lucky enough to obtain an academic job, gender biases can influence the relative resources allocated to faculty, as Nancy Hopkins discovered when she and a senior faculty committee studied this problem at MIT. The data were so convincing that MIT president Charles Vest publicly admitted that discrimination was responsible. For talented women, academia is all too often not a meritocracy.

"I am suspicious when those who are at an advantage proclaim that a disadvantaged group of people is innately less able."

In denial

Despite these studies, very few men or women are willing to admit that discrimination is a serious problem in science. How is that possible? Valian suggests that we all have a strong desire to believe that the world is fair[11]. Remarkably, women are as likely as men to deny the existence of gender-based bias[13]. Accomplished women who manage to make it to the top may 'pull up the ladder behind them', perversely believing that if other women are less successful, then one's own success seems even greater. Another explanation is a phenomenon known as 'denial of personal disadvantage', in which women compare their advancement with other women rather than with men[11].

My own denial of the situation persisted until last year, when, at the age of 50, several events opened my eyes to the barriers that women and minorities still face in academia. In addition to the Summers speech, the National Institutes of Health (NIH) began the most prestigious competition they have ever run, the Pioneer Award, but with a nomination process that favoured male applicants[14]. To their credit, in response to concerns that 60 of 64 judges and all 9 winners were men, the NIH has revamped their Pioneer Award selection process to make it fairer. I hope that the Howard Hughes Medical Institute (HHMI) will address similar problems with their investigator competitions. When it comes to bias, it seems that the desire to believe in a meritocracy is so powerful that until a person has experienced sufficient career-harming bias themselves they simply do not believe it exists.

My main purpose in writing this commentary is that I would like female students to feel that they will have equal opportunity in their scientific careers. Until intolerance is addressed, women will continue to advance only slowly. Of course, this feeling is also deeply personal to me (see 'Personal experiences'). The comments of Summers, Mansfield, Pinker and Lawrence about women's lesser innate abilities are all wrongful and personal attacks on my character and capabilities,

P. TURNLEY/CORBIS

Few women, as well as men, are willing to admit that there is discrimination in academia.

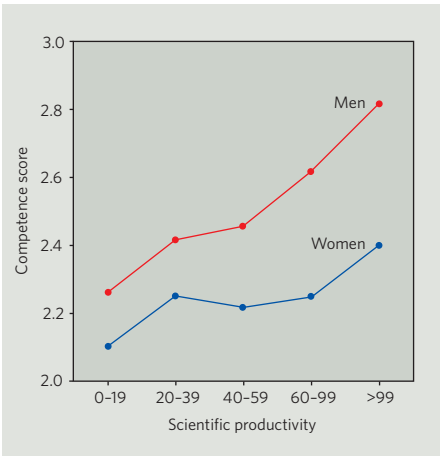


Figure 2 | Competence scores awarded after peer review. Peer reviewers in Sweden award lower competence scores to female scientists than to similarly productive male scientists.

©2006 Nature Publishing Group

PERSONAL EXPERIENCES

As a transgendered person, no one understands more deeply than I do that there are innate differences between men and women. I suspect that my transgendered identity was caused by fetal exposure to high doses of a testosterone-like drug. But there is no evidence that sexually dimorphic brain wiring is at all relevant to the abilities needed to be successful in a chosen academic career. I underwent intensive cognitive testing before and after starting testosterone treatment about 10 years ago. This showed that my spatial abilities have increased as a consequence of taking testosterone. Alas, it has been to no avail; I still get lost all the time when driving (although I am no longer willing to ask for directions). There was one innate difference that I was surprised to learn is apparently under direct control of testosterone in adults — the ability to cry easily, which I largely lost upon starting hormone treatment. Likewise, male-to-female transgendered individuals gain the ability to cry more readily. By far, the main difference that I have noticed is that people who don't know I am transgendered treat me with much more respect: I can even complete a whole sentence without being interrupted by a man.

as well as on my colleagues' and students' abilities and self esteem. I will certainly not sit around silently and endure them.

Mansfield and others claim that women are more emotional than men. There is absolutely no science to support this contention. On the contrary, it is men that commit the most violent crimes in anger — for example, 25 times more murders than women. The only hysteria that exceeded MIT professor Nancy Hopkins' (well-founded) outrage after Larry Summers' comments was the shockingly vicious news coverage by male reporters and commentators. Hopkins also received hundreds of hateful and even pornographic messages, nearly all from men, that were all highly emotional.

Taboo or untrue?

There is no scientific support, either, for the contention that women are innately less competitive (although I believe powerful curiosity and the drive to create sustain most scientists far more than the love of competition). However, many girls are discouraged from sports for fear of being labelled tomboys. A 2002 study did find a gender gap in competitiveness in financial tournaments, but the authors suggested that this was due to differences in self confidence rather than ability[15]. Indeed, again and again, self confidence has been pointed to as a factor influencing why women 'choose' to leave science and engineering programmes. When women are repeatedly told they are less good, their self confidence falls and their ambitions dim[16]. This is why Valian has concluded that simply raising expectations for women in science may be the single most important factor in helping them make it to the top[11].

Steven Pinker has responded to critics of the Larry Summers Hypothesis by suggesting that they are angry because they feel the idea that women are innately inferior is so dangerous that it is sinful even to think about it[17]. Harvard Law School professor Alan Dershowitz sympathizes so strongly with this view that he plans to teach a course next year called 'Taboo'. At Harvard we must have veritas; all ideas are fair game. I completely agree. I welcome any future studies that will provide a better understanding of why women and minorities are not advancing at the expected rate in science and so many other professions.

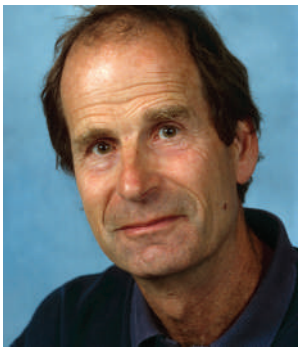
But it is not the idea alone that has sparked anger. Disadvantaged people are wondering why privileged people are brushing the truth under the carpet. If a famous scientist or a president of a prestigious university is going to pronounce in public that women are likely to be innately inferior, would it be too much to ask that they be aware of the relevant data? It would seem that just as the bar goes way up for women applicants in academic selection processes, it goes way down when men are evaluating the evidence for why women are not advancing in science. That is why women are angry. It is incumbent upon those proclaiming gender differences in abilities to rigorously address whether suspected differences are real before suggesting that a whole group of people is innately wired to fail.

What happens at Harvard and other universities serves as a model for many other institutions, so it would be good to get it right. To anyone who is upset at the thought that free speech is not fully protected on university campuses, I would like to ask, as did third-year Harvard Law student Tammy Pettinato: what is the difference between a faculty member calling their African-American students lazy and one pronouncing that women are innately inferior? Some have suggested that those who are angry at Larry Summers' comments should simply fight words with more words (hence this essay). In my view, when faculty tell their students that they are innately inferior based on race, religion, gender or sexual orientation, they are crossing a line that should not be crossed — the line that divides free speech from verbal violence — and it should not be tolerated at Harvard or anywhere else. In a culture where women's abilities are not respected, women cannot effectively learn, advance, lead or participate in society in a fulfilling way.

Take action

Although I have argued that the Larry Summers Hypothesis is incorrect and harmful, the academic community is one of the most tolerant around. But, as tolerant as academics are, we are still human beings influenced by our culture. Comments by Summers and others have made it clear that discrimination remains an under-recognized problem that is far from solved. The progress of science increasingly depends on the global community, but only 10% of the world's population is male and caucasian. To paraphrase Martin Luther King, a first-class scientific enterprise cannot be built upon a foundation of second-class citizens. If women and minorities are to achieve their full potential, all of us need to be far more proactive. So what can be done?

First, enhance leadership diversity in academic and scientific institutions. Diversity provides a substantially broader point of view, with more sensitivity and respect for different perspectives, which is invaluable to any organization. More female leadership is vital in lessening



H. BORDEN; R. FRIEDMAN/CORBIS; MRC CAMBRIDGE

Stephen Pinker, Larry Summers and Peter Lawrence (left to right) all argue that innate differences are at least partly to blame for the failure of women to advance in science.

©2006 Nature Publishing Group

the hostile working environment that young women scientists often encounter. In addition to women and under-represented minority groups, we must not forget Asians and lesbian, gay, bisexual and transgendered folks. There are enough outstanding scientific leaders in these racial and gender groups that anyone with a will to achieve a diverse leadership in their organization could easily attain it.

Speak out

Second, the importance of diverse faculty role models cannot be overstated. There is much talk about equal opportunity, but, in practice, serious attention still needs to be directed at how to run fair job searches. Open searches often seem to be bypassed entirely for top leadership positions, just when it matters most — search committees should not always be chaired by men and the committee itself should be highly diverse[14,18]. Implementation of special hiring strategies and strong deans willing to push department chairs to recruit top women scientists are especially effective. It is crucial in the promotion process that merit be decided by the quality, not quantity, of papers published.

Women faculty, in particular, need help from their institutions in balancing career and family responsibilities. In an increasingly competitive environment, women with children must be able to compete for funding and thrive. Why can't young faculty have the option of using their tuition benefits, in which some universities pay part of the college tuition fees for the children of faculty, for day care instead? Tuition benefits will be of no help if female scientists don't make tenure. And institutions that have the financial capability, such as HHMI, could help by making more career-transition fellowships available for talented women scientists.

> "Simply raising expectations for women in science may be the single most important factor in helping them make it to the top."
> — Virginia Valian

Third, there should be less silence in the face of discrimination. Academic leadership has a particular responsibility to speak out, but we all share this responsibility. It takes minimal effort to send a brief message to the relevant authority when you note a lack of diversity in an organization or an act of discrimination. I don't know why more women don't speak out about sexism at their institutions, but I do know that they are often reluctant, even when they have the security of a tenured faculty position. Nancy Hopkins is an admirable role model, and it is time that others share the burden. It doesn't only have to be women that support women. I was deeply touched by the eloquent words of Greg Petsko[19] following Summers' comments. And it has been 30 years since I was a medical student, but I still recall with gratitude the young male student who immediately complained to a professor who had shown a slide of a nude pin-up in his anatomy lecture.

J. WEST/ALAMY

At school, girls and boys show similar levels of ability in the sciences.

Fourth, enhance fairness in competitive selection processes. Because of evaluation bias, women and minorities are at a profound disadvantage in such competitive selection unless the processes are properly designed[11,12,14,18]. As the revamped NIH Pioneer Award demonstrates, a few small changes can make a significant difference in outcome. By simply changing the procedure so that anyone can self-nominate and by ensuring a highly diverse selection committee, the number of women and minority winners went up to more than 50% from zero. This lesson can and should now be applied to other similar processes for scientific awards, grants and faculty positions. Alas, too many selection committees still show a striking lack of diversity — with typically greater than 90% white males. When selection processes are run fairly, reverse discrimination is not needed to attain a fair outcome.

Confidence booster

Finally, we can teach young scientists how to survive in a prejudiced world. Self-confidence is crucial in advancing and enjoying a research career. From an early age, girls receive messages that they are not good enough to do science subjects or will be less liked if they are good at them. The messages come from many sources, including parents, friends, fellow students and, alas, teachers. When teachers have lower expectations of them, students do less well. But we are all at fault for sending these messages and for remaining silent when we encounter them. Teachers need to provide much more encouragement to young people, regardless of sex, at all stages of training.Occasional words of encouragement can have enormous effects.

All students, male and female, would benefit from training in how to be more skillful presenters, to exert a presence at meetings by asking questions, to make connections with faculty members who may help them to obtain grants and a job, and to have the leadership skills necessary to survive and advance in academia. Because women and minorities tend to be less confident in these areas, their mentors in particular need to encourage them to be more proactive. I vividly recall my PhD supervisor coming with me to the talks of famous scientists and forcing me to introduce myself and to ask them questions. There is a great deal of hallway mentoring that goes on for young men that I am not sure many women and minorities receive (I wish that someone had mentioned to me when I was younger that life, even in science, is a popularity contest — a message that Larry Summers might have fo und helpful as well). It is incumbent on all of us who are senior faculty to keep a look out for highly talented young people, including women and minority students, and help them in whatever way possible with their careers. ■

Ben A. Barres is at Stanford University School of Medicine, Department of Neurobiology, Fairchild Building Room D235, 299 Campus Drive, Stanford, CA 94305-5125, USA.
e-mail: barres@stanford.edu

1. Summers, L. *Letter to the Faculty Regarding NBER Remarks* www.president.harvard.edu/speeches/summers/2005/facletter.html (2005).
2. *The Science of Gender and Science. Pinker vs. Spelke: A Debate* www.edge.org/3rd_culture/debate05/debate05_index.html (2005).
3. Lawrence, P. A. *PLoS Biol.* **4,** 13–15 (2006).
4. Baron-Cohen, S. *The Essential Difference: Men, Women, and the Extreme Male Brain* (Allen Lane, London, 2003).
5. Mansfield, H. *Manliness* (Yale Univ. Press, New Haven, 2006).
6. Gould, S. J. *The Mismeasure of Man* (W. W. Norton & Co, New York, 1996).
7. Steele, C. M. *Am. Psychol.* **52,** 613–629 (1997).
8. Spelke, E. S. *Am. Psychol.* **60,** 950–958 (2005).
9. Leahey, E. & Guo, G. *Soc. Forces* **80.2,** 713–732 (2001).
10. Xie, Y. & Shauman, K. *Women in Science: Career Processes and Outcomes* (Harvard Univ. Press, Cambridge, 2003).
11. Valian, V. *Why So Slow?* (MIT Press, Cambridge, 1998).
12. Wennerås, C. & Wold, A. *Nature* **387,** 341–343 (1997).
13. Rhode, D. L. *Speaking of Sex: The Denial of Gender Inequality* (Harvard Univ. Press, Cambridge, 1997).
14. Carnes, M. *et al. J. Womens Health* **14,** 684–691 (2005).
15. Gneezy, U., Niederle, M. & Rustichini, A. *Q. J. Econ.* **18,** 1049–1074 (2003).
16. Fels, A. *Necessary Dreams* (Pantheon Press, New York, 2004).
17. Pinker, S. *New Repub.*15 (14 Feb, 2005).
18. Moody, J. *Faculty Diversity: Problems and Solutions* (Taylor and Francis, New York, 2004).
19. Petsko, G. A. *Genome Biol.* **6,** 1–3 (2005).

©2006 Nature Publishing Group

BOOKS & ARTS

God is bred

Religious belief can be viewed as an adaptation that was favoured as the human brain evolved.

Six Impossible Things Before Breakfast: The Evolutionary Origins of Belief
by Lewis Wolpert
Faber and Faber: 2006. 243 pp. £14.99
To be published by WW Norton in the United States in January 2007

Crispin Tickell

There are, of course, many more than six impossible things to believe in before breakfast (as Alice said in *Through The Looking Glass*). The range of human beliefs — past, present and probably future — is almost infinite. This book goes back beyond beliefs to trace the origin of belief as such, and its role in the evolution of human brains and behaviour.

Lewis Wolpert's central thesis is simple. At the beginning are the brain and sense organs, which together help organisms decide how to move. The next step, a long way along the track, is the use of tools, an ability humans share with other species, such as some birds and other primates. In the human case this leads gradually to understanding cause and effect. By then the human brain has become much larger. There may have been what Wolpert calls a "magic moment" when a proto-language acquired syntax and became the kind of language we know. From that comes the evolution of "a belief-generating machine, an engine that can produce beliefs with little relation to what is actually true". Whether its products are true or not, the machine serves a variety of useful purposes, and has a strong genetic element deriving from natural selection: the 'God gene'. It is now an essential part of the human condition.

How this all happened is a fascinating story, well told here. Not everyone will agree with each stage. Wolpert proceeds through observations on animal behaviour, child psychology, and the archaeological evidence for increasing tool complexity, with language to match. The transition from understanding cause and effect to systems of belief is difficult to explain but obviously crucial. If we look at our own mental processes, we can see the way in which we begin by accepting authority as part of growing up, and even if we subsequently change our minds, we try to maintain a consistent view, either with or without contradictions. All too easily we misinterpret coincidences, accept anecdotal evidence and indulge in circular reasoning. New ideas are often painful, and are strenuously resisted. Rarely do we embark on a comprehensive view of the assumptions that govern our daily lives.

T. VAN HOUTRYVE/PANOS

Prayer matter: religion and paranormal beliefs provide comfort, explanation and meaning.

Once a belief is established, it tends to perpetuate itself. Wolpert discusses in somewhat hostile fashion the role of what Richard Dawkins has called memes, or units of information that, unlike genes, can pass laterally across cultures rather than vertically by inheritance. But he rightly brings out our ability to deceive ourselves, and this brings him to a tactful analysis of religion in its many forms, and a less tactful analysis of paranormal beliefs. Both can serve useful social purposes: they can supply explanations where none is otherwise forthcoming, provide consolation to the unhappy or bereaved, promote group solidarity and cohesion, and give overall purpose to life. They can also do a lot of harm: the history of religious intolerance and fanaticism is an alarming reminder of what people can do to each other, from warfare to the burning of witches, and some of this is still with us today.

Why we believe as we do also raises questions of health. Poor understanding of cause and effect has bedevilled medical treatment down the ages. Only recently have we introduced a more scientific approach, and this is far from universally accepted. Modern humans seem particularly vulnerable to schizophrenia, depression and other ailments in which delusions play a major part. The will of God is still seen as critical in many societies. Indeed, in our own society many reject reductionist interpretations of disease, and rightly or wrongly go for alternative treatments based more on faith than on anything else.

What is the answer? Is there one? Wolpert firmly goes for science. Although science can be counterintuitive (he refers to "the unnatural nature of science"), it provides the only fundamental explanation of how the world works. It is in constant evolution as knowledge accumulates, it is self-correcting, and it is universally valid without cultural bias. It is much more than the kind of relative social construct suggested by some sociologists, and if we need a basis for belief, it is the best available. As Wolpert concludes, we have to respect the beliefs of others, but it is their — and our — actions that ultimately matter.

This is an admirable short guide to an immensely complex subject. I could have wished for a little less technical jargon — "confabulation" and "heuristics" — and sometimes the text gets lost in quotations from others. It is unfortunate that this book should have come out at the same time as Daniel Dennett's *Breaking The Spell* (Viking/Allen Lane, 2006) on broadly the same subject (see *Nature* **439,** 535; 2006 for a review). But Wolpert's book is more succinct, and much better argued, and I would go for it every time. ■

Crispin Tickell is director of the Policy Foresight Programme in the James Martin Institute for Science and Civilization at Oxford University, Oxford OX1 1HP, UK.

©2006 Nature Publishing Group

NATURE|Vol 442|13 July 2006

Keep it in the family

Genes and Behavior: Nature–Nurture Interplay Explained
by Michael Rutter
Blackwell: 2006. 272 pp. $24.95, £14.99

Svenn Torgersen

Today, every generally well-informed person is familiar with the findings of genetic research. Not only do we read that genes have a big influence on the development of personality, intelligence, diseases and disorders, but we have even learned that political and religious orientations, such as being a hawk or a dove, are to a high degree influenced by our genes. Outstanding genetic researchers tell us that our family background is not important; it is our genes and our personal private experiences, friends and acquaintances that count. And discoveries of new genes 'for' various behaviours and disorders are proclaimed daily.

Michael Rutter, the renowned child psychiatrist, challenges many of these views in his new book, *Genes and Behavior*. Modestly, he announces "I am not trained as a geneticist but I have been a user of behavioral genetics for some 30 years and I have made myself informed about genetic mechanisms and genetic issues." The book leaves no doubt that he is informed and he gives an excellent popular introduction to genetics, including its most modern variants. He describes how the functional, biological effects of genes (gene expression) are influenced by factors outside the gene — both environmental conditions and other genes. He covers well the two topics of gene–environment correlation and interaction. The first arises as we select, shape and evoke reactions from the environment according to our genes, whereas the second deals with our genetic susceptibility to react to particular types of environment, or to put it another way: the effects of certain genes depend on the environment.

In addition, the book seeks to challenge the claims of those whom Rutter labels as "environmental evangelists" and "genetic evangelists". Environmental evangelists disregard the fact that genes influence behaviour, whereas genetic evangelists go in for hyperbolic claims about genetic research and the role of genes.

In Rutter's view, a lot of the findings in quantitative behavioural genetics are either mistaken or misleading, and those that are correct are not interesting. He maintains that all behaviours and disorders are influenced by genes, and that attempts to quantify their relative contribution are unimportant. He considers such findings to be at best a source of inspiration for hypotheses concerning what he thinks is really important — finding the causes of behaviour and its disorders through the study of gene–environment interactions and correlations.

Few in the field will agree with this conclusion. The numbers do matter. For example, if a disorder has a heritability of 80% and one identical twin has it, the other twin is 40 times more likely than a randomly selected unrelated person to have it too, assuming a prevalence of 1% for the condition in the population at large. In contrast, if a disorder has a heritability of 20%, the identical twin is only three times more likely to get the disorder. Rutter's additional objection, however, that the magnitude of genetic effects depends on culture and time, is widely accepted in the field, and has inspired researchers to study culturally different samples, and even to include different time periods in their study designs.

A window on personality: identical twins end up similar even when reared apart.

HULTON ARCHIVE/GETTY

Even worse in Rutter's opinion is the contention that the family in which you grow up does not matter much — within the normal range — for your personality development and mental health. Here, Rutter's viewpoint corresponds to common sense. Yet all researchers who have interviewed several sets of twins are struck by how different fraternal twins can be and how similar identical twins are.

One could concede that this effect is the result of gene–environment interaction, making genetically different fraternal twins react differently to the same family environment. Nevertheless, if family environment is so important, as Rutter believes it is, how do we then explain the findings that identical twins who are reared apart are almost as similar to each other as identical twins reared together? And why is there almost no similarity between adoptive relatives, whereas adopted-away children are almost as similar to their biological mothers as if they had grown up with them? Rutter rebuffs this research on methodological grounds, holding that the samples are not representative, and even suggesting that many of the studies have not been published in peer-reviewed journals. A simple literature search would have revealed that more than a few studies of twins reared apart have appeared in peer-reviewed publications.

As one might expect, Rutter opposes claims that there is a gene 'for' a specific behaviour or disorder. Even when he accepts some of the results from studies that have detected such relationships, his opinion is that the pathway from the gene to actual behaviour is too long and complicated to make the connection definitive. Hence, the relationship between gene and environment is probabilistic and not deterministic, modified both by our will and — ultimately — by the environmental conditions that interact with the genes. Discounting those adoption studies showing the limitations of gene–environment interaction, he focuses attention on those that do demonstrate such an interaction.

Another prominent theme in the book is gene–environment correlations: that is, how genes come to shape our lives for better or worse. Rutter gives a thorough introduction to the state of the art in this field. This research is in accordance with Rutter's strong interest in social issues and the conditions of people's lives. And, of course, the study of gene–environment correlations also forms an important part of Rutter's quest to find the causes of behavioural disorders.

If you want an inspiring contribution to the debate in this highly topical area of research and also want to learn about the most up-to-date approaches to genetic research, then this is the book to choose.

In the early 1970s, Rutter became famous for challenging environmental determinism in his book, *Maternal Deprivation Reassessed* (Penguin, 1972). Now, more than 30 years later when, in his view, the opposite *zeitgeist* is about to prevail, he challenges the genetic determinists. He has closed the circle. ■

Svenn Torgersen is in the Department of Psychology, Oslo University, Postbox 1074, Blindern, N-0317 Oslo, Norway.

©2006 Nature Publishing Group

Throwaway culture

Made to Break: Technology and Obsolescence in America
by Giles Slade
Harvard University Press: 2006. 336 pp. $27.95, £18.95, €25.80

John Emsley

The subtitle of Giles Slade's *Made to Break* suggests that the book is an update of Vance Packard's *The Hidden Persuaders* (Longmans, 1957), which I remember reading as a student in the 1960s. At the time I was surprised and shocked to discover that US industry was deliberately producing products with a short lifespan. Planned obsolescence it was called, but little did I realize that this was to be the secret weapon that won the Cold War for the United States. That little-known story is one of the intriguing *Made to Break*'s hidden gems.

In the 1970s and 1980s the Soviet Union was falling behind the United States in the manufacture of microchips and computers, yet it needed them if it was to exploit and export its vast reserves of oil and gas to earn desperately needed foreign currency. A US embargo was denying them the necessary technology, so Soviet agents resorted to covert methods to obtain it, while the CIA did all it could to thwart them. Then Gus Weiss, an economist and government security adviser, made an ingenious suggestion: allow the Soviet agents to succeed, but make sure they bought specially doctored microchips that looked like the real thing but had inbuilt obsolescence.

This inbuilt obsolescence included sudden catastrophic malfunctioning, such as suddenly instructing a pump to work at a pressure far higher than a pipeline could withstand. The result was a series of spectacular explosions, some so large that they were observed by satellites. One failure burst an oil pipeline, creating a lake more than 10 kilometres long and 2 metres deep before it was brought under control. This industrial sabotage ensured that the Soviet Union lost the cold war says Slade, and he makes it almost believable.

The phrase 'planned obsolescence' was first used by Bernard London in 1932. He proposed it as a possible solution to the Depression, which was being made worse because people had stopped buying. Planned obsolescence looked like the solution to economic stagnation in a world with abundant natural resources and unemployed labour. London's suggestion was at variance with the idea of the father of mass production, Henry Ford, who assumed that people would obviously want cars to last — and a 15 million Model Ts had shown that the idea had its appeal.

In fact many affluent people wanted an excuse to buy a new car, and by realizing this and pandering to it, General Motors (GM) was to become the world's largest company. In the late 1920s they began to boost sales by launching a new model every year, even if it was merely a style change. It worked — but it went too far. By the late 1950s, GM cars had become chrome-finned monstrosities and the US public went to the other extreme, beginning to buy small cars, such as the Volkswagen Beetle, that were built to last. Planned obsolescence had become self-defeating.

D. MASON/CORBIS

Booted out: technological progress has largely replaced planned obsolence as a spur to consumerism.

Made to Break is not just about planned obsolescence; indeed, it is more about the unplanned obsolescence that technological development brings. Slade tells the tale of the early days of FM radio and how the emerging television industry commandeered its broadcast frequencies, making all FM radios of the 1940s obsolescent. He also has a chapter on the introduction of nylon, but he misses a chance to explain how the hosiery industry, faced with apparently everlasting stockings and tights, introduced obsolescence by making the thread so fine it broke easily. Today we have technological obsolescence on a large scale — witness the dumping of PCs in favour of laptops — and the book ends with a chapter on mobile phones, where technological obsolescence is such that some users upgrade every year.

Made to Break is both entertaining and thought-provoking, but finally somewhat unsatisfying in that it does not say whether planned obsolescence is still with us. I suspect it disappeared 20 years ago because it was no longer needed. Perhaps it will re-emerge this century if the new method that keeps us spending — unlimited credit — fails. For a decade now that is what has kept the wheels of industry turning, at least in China. ■

John Emsley is in the Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. He is the author of several books including *Vanity, Vitality, and Virility* and *Elements of Murder*.

The politics of space

The Last of the Great Observatories: Spitzer and the Era of Faster, Better, Cheaper
by George Rieke
University of Arizona Press: 2006. 264 pp. $40 (hbk), $19.95 (pbk)

Steven Beckwith

Modern space observatories, in at least one respect, are like ancient monuments: planned by elders and built by their children for the benefit of their grandchildren. The creation of the Spitzer Space Telescope took 20 years from start to launch, encompassing nearly half of a scientific career. George Rieke, principal investigator on one of three Spitzer instruments, worked on the project from its inception in 1983 through to its launch in 2003. *The Last of the Great Observatories* chronicles the evolution of the project, its highs and lows, and the many near-death experiences due to political, management and technical problems. He also describes the emotional roller-coaster that a scientist whose career is wedded to one project's success rides along the way. For anyone who has taken part in a large space project or who wants to appreciate the commitment required to do so, this is a good read.

Rieke goes beyond telling a good story and tries to glean lessons for future projects from the Spitzer experience. Spitzer's 20-year genesis spanned several eras of NASA management styles. The book spends considerable time discussing the contrast between the Apollo-era command-and-control organization adopted from the military and NASA administrator Dan Goldin's "better, faster, and cheaper" era, as well as the individual-tinkering style appropriate to small university groups.

Rieke shows the difficulty of coordinating large teams from disparate cultures: scientists, whose rewards depend on individual achievement; contractors, who focus on fees and developing technical prowess; and managers,

©2006 Nature Publishing Group

who worry most about time, money and team performance. Rieke's own preferences for working style are barely disguised in his writing, and his partisanship is on display when he compares Spitzer to other great observatories or discusses the contributions of different teams in the project, a subtle reminder of how difficult it was to keep the Spitzer team working together. He nevertheless manages to show how each style has its place, and acknowledges that each has strengths and weaknesses in coping with the different scales and complexities of NASA projects. Many thoughtful insights are unfortunately relegated to an appendix that might have made a nice final chapter.

The issues tackled in this book are not unique to space science — almost all large projects, from mediaeval cathedrals to the Brooklyn Bridge, required more time and money than their promoters believed at the outset. But the modern era poses unique challenges, with projects that depend more on the coordination of people's knowledge than on marshalling their physical labour. These challenges are nicely chronicled in Tracy Kidder's *The Soul of a New Machine* (Little Brown, 1981), a Pulitzer-prizewinning book about a team of engineers at Data General Corporation who created a new computer. They will be increasingly important for the future health of space science. Despite NASA's enormous reservoir of talented and well-intentioned managers, large missions and their inevitable cost overruns put pressure on the rest of space science.

The Last of the Great Observatories recognizes that although the Spitzer experience is not applicable to all space projects, the human emotions involved in the organization of these complex undertakings may well be universal. The description of these emotions will be a valuable guide to scientists and engineers ambitious enough to participate in large-scale, multi-decade projects. ■

Steven Beckwith is past director of the Space Telescope Science Institute in Baltimore, Maryland, and a professor of physics and astronomy at Johns Hopkins University.

SCIENCE IN CULTURE

The medium is the message

PowerPoint presentations and the culture of pitch.

Martin Kemp

Most scientists live in a world densely populated with graphic representations of data. The 15 June issue of *Nature* contained more than 70 graphs or graph-like presentations of results, to say nothing of a clutch of advertisements that incorporate such devices. Every graph or chart embodies analytical assumptions, and at a deeper level the implicit cognitive structures that shape every phase of a scientific study.

A graph is not only a crucial tool in plotting data within parameters that are taken as significant, but is also an integral part of the style of a presentation. It serves the rhetoric of irrefutable precision that operates whenever we aspire to proceed in a 'scientific' manner.

These remarks are occasioned by a coruscating attack upon PowerPoint presentations in Edward Tufte's latest book, *Beautiful Evidence* (Graphics Press, 2006). Tufte is an expert in public affairs at Yale University; he is also an expert in the presentation of statistical evidence and a sculptor, and has established himself as the world's leading analyst of graphic information.

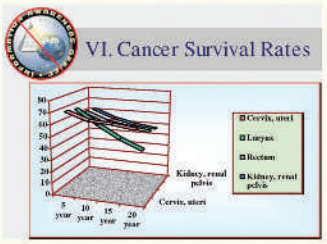
Centring his attack on the revelations in the 2005 report on the 2003 Columbia space-shuttle disaster, Tufte lays bare the special kinds of hierarchies, abbreviations, neologisms, linguistic tropes and implicit cognitive structures that were embedded in the PowerPoint presentations used by NASA officials. While the damaged craft wheeled in orbit, their modes of presentation worked against a proper analysis of the complex uncertainties and risks occasioned by the foam debris that had struck the wing during take-off.

Echoing the criticisms in the report itself, and comparable strictures expressed by Richard Feynman at the inquiry into the 1986 Challenger disaster, Tufte concludes that "the pitch-style typography of PowerPoint is hopeless for science and engineering". The rat-a-tat style of bullet points, arranged in aggressive hierarchies and coupled with compressed jargon, insider acronyms and summary graphics, took over from the need to present supporting data, documentation and modes of analysis.

Rather than being narrative reports, capable of debating the complexities and laying them open for decision-makers, the NASA presentations relied on a series of PowerPoint slides, each typically containing 40 or so words (equivalent to a mere 8 seconds of normal reading). Typically, each slide is succeeded by the next in such a way that vital logical continuities fall between the gaps.

Cancer survival (estimated relative rates, %)

	5 year	10 year	15 year	20 year
Prostate	98.8	95.2	87.1	81.1
Thyroid	96.0	95.8	94.0	95.4
Testis	94.7	94.0	91.1	88.2
Melanomas	89.0	86.7	83.5	82.8
Breast	86.4	78.3	71.3	65.0
Hodgkin's	85.1	79.8	73.8	67.1
Uterus	84.3	83.2	80.8	79.2
Urinary, bladder	82.1	76.2	70.3	67.9
Cervix, uteri	70.5	64.1	62.8	60.0
Larynx	68.8	56.7	45.8	37.8
Rectum	62.6	55.2	51.8	49.2
Kidney, renal	61.8	54.4	49.8	47.3
Colon	61.7	55.4	53.9	52.3
Non-Hodgkin's	57.8	46.3	38.3	34.3
Oral cavity	56.7	44.2	37.5	33.0
Ovary	55.0	49.3	49.9	49.6
Leukemia	42.5	32.4	29.7	26.2
Brain	32.0	29.2	27.6	26.1
Myeloma	29.5	12.7	7.0	4.8
Stomach	23.8	19.4	19.0	14.9
Lung, bronchus	15.0	10.6	8.1	6.5
Esophagus	14.2	7.9	7.7	5.4
Liver, bile duct	7.5	5.8	6.3	7.6
Pancreas	4.0	3.0	2.7	2.7



E. TUFTE

A typical template in PowerPoint is the standard line graph, such as the one of cancer survival rates pictured left. It is linked directly to a data sheet and can be adapted for any purpose in which such a display is desired. The three-dimensionality of the plot increases its graphic impact and enhances its air of reality. But we do not necessarily know anything about the nature of the data sets that lie behind it, nor of the methods of data gathering and modes of analysis. A 'result' is all too readily presented as 'significant' without meeting the statistical criteria for significance.

Such charts, delivered with pleasing facility by software packages, permeate business and other non-scientific subjects that generate quantitative information. Government and other official reports, and articles in the press on economics and the social sciences are permeated by facile graphics that look all too convincing. Of course, real science never resorts to such sloppy shorthand — does it?

Let us leave the last word to Tufte: "Making an evidence presentation is a moral act as well as an intellectual activity. To maintain standards of quality, relevance, and integrity of evidence, consumers of presentations should insist that presenters be held intellectually and ethically responsible for what they show and tell. Thus consuming is also an intellectual and moral activity."

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK.

©2006 Nature Publishing Group

NEWS & VIEWS

NEUROSCIENCE

Converting thoughts into action

Stephen H. Scott

There is a clear need to help people who have brain or spinal-cord damage to communicate and interact with the outside world. Progress to that end is being made with brain-implantation technology.

Using our thoughts to control a computer or robot used to be the realm of science-fiction writers. But scientists have been making concerted efforts to develop the technology required to convert brain signals into commands, to support communication, mobility and independence for paralysed people. In this issue, two papers shift the notion of such 'implantable neuromotor prosthetics' from science fiction towards reality. Hochberg *et al.* (page 164)[1] describe the first implantation of electrode arrays into the brain of a paralysed man; these allowed him to use his thoughts (motor intentions) to directly control devices such as a computer mouse. In addition, Santhanam and colleagues (page 195)[2] describe a new software approach for extracting intended actions from the neural activity in the brain of monkeys that dramatically improves the potential speed and performance of implantable neuromotor prosthetics.

Although it may someday be possible to reconnect damaged neural pathways by directing the regrowth of neurons, neuroprosthetics provide another potential approach to permit individuals with severe neurological injuries to interact with the environment. Damage to the nervous system means these individuals lose their motor control — that is, they can no longer directly control their muscles. Neuroprosthetics aim to bypass this damage by recording brain activity that reflects the individual's motor intentions. These signals are then used either to reanimate paralysed muscles using electrical stimulation, or to control physical devices directly, such as artificial limbs, computer cursors or wheelchairs. Several different approaches have been developed ranging from non-implantable technologies that record electroencephalographic (EEG) activity using removable electrodes placed on the scalp surface[3], to implantable devices that use microelectrodes to detect the activities of individual neurons[4–7].

Implantable neuromotor prosthetics build on basic research looking at the neural basis of the planning and control of movement, predominantly in monkeys. A brain region of particular interest for neuroprosthetic research is the primary motor cortex, a major region involved in controlling voluntary movements. Others include the premotor and posterior parietal cortex, which are principally involved in planning movements, providing instructions for the motor cortex to then act on. In monkeys, the activities of large numbers of neurons can be recorded simultaneously and be used to predict motor intention[4] and limb motion[5], or to move cursors on a computer screen[6,7]. Animal experimentation remains essential for testing and developing implantable neuroprosthetics, but obviously at some point the great leap into human patients needs to be made.

To this end, Hochberg *et al.*[1] recruited a man who can no longer move his limbs because his spinal cord is completely severed. They implanted an array of tiny electrodes into his primary motor cortex, and tested whether the activity of neurons recorded there could control prosthetic devices. Remarkably, the implanted electrodes allowed the patient to control a computer cursor and rudimentary movement of robotic devices (the associated videos are online in Supplementary Information[8]). This is not the first neuromotor prosthetic that has been implanted into a person — a previous experiment used a couple of implanted electrodes to generate limited horizontal control of a cursor[9]. But this study reports several significant advances.

First, even though the patient had been paralysed three years earlier, the neural activity in his primary motor cortex seemed relatively

©2006 Nature Publishing Group

NATURE|Vol 442|13 July 2006

normal. The neurons were spontaneously active and the subject could modulate this activity based on instructions such as, "Move your arm". This shows that activity in this region can still be modulated by a subject's motor intentions, even though axons projecting from this region to the spinal cord have been severed and the brain region has not been used to control limb movements for several years.

Second, the calibration of the device was achieved by simply asking the subject to imagine moving his hand to track a moving cursor on the computer screen. This process took only minutes, much less than the weeks or months of training required for current non-invasive EEG systems.

Third, after this calibration, the subject was immediately able to perform, to some degree, other computer-based tasks including reading e-mail and playing simple computer games. He could also control a television and even simple robotic devices. This flexibility in the use of neural signals is notable because the user of a similar neuroprosthetic will potentially be able to control a range of devices from computers to wheelchairs.

This research suggests that implanted prosthetics are a viable approach for assisting severely impaired individuals to communicate and interact with the environment. But there have also been substantial advances in non-invasive approaches to recording brain activity that allow comparable control of cursor motion[3]. So why use an invasive technology, with all the risks related to surgery and the long-term care inherent in implantable devices, when non-invasive technologies are available?

Several differences between the neural signals recorded from invasive and non-invasive electrodes suggest that implantable prosthetics will have greater long-term potential. EEG recordings reflect the averaged activity of millions of neurons, creating a signal with limited spatial and temporal resolution. By contrast, direct recordings using hundreds of microelectrodes, each monitoring the activity of a single or a small number of neurons, can create highly complex control signals.

The potential speed of implantable neuromotor prosthetics is demonstrated by Santhanam *et al.*[2]. A common strategy for the software algorithms used to convert neural recordings into a usable output (including the one used by Hochberg *et al.*[1]) is to translate neural activity into an estimate of hand trajectory or cursor motion. Santhanam *et al.* simplified the problem by using neural activity from the premotor cortex of monkeys to predict the location of spatial targets. By optimizing for the number of potential targets and the time period used to sample neural activities, they were able to extract the correct spatial location in about 250 ms from about 100 single and multi-unit recordings. This performance corresponds to retrieving about 6.5 bits of information per second — which would allow a person to type at approximately 15 words per minute. This is substantially faster than can be done using existing implantable or non-implantable technologies.

The algorithm developed by Santhanam and colleagues is not directly applicable to neuroprosthetics because it relies on the animal's learned response to a visual stimulus. However, this approach should be easily modified for humans and be driven not by sensory stimuli, but by the patient's own intentions, much as one converts a word into a series of key presses on a computer keyboard.

Many considerable problems must be addressed before this technology can be put to regular clinical use. It is not clear for how long the very fine microelectrodes can be used to record neural activity. Patients with spinal-cord injuries are often young and would need to use these technologies for many decades. Moreover, the device used by Hochberg *et al.* involved passing a large bundle of wires directly through the skin to a connector attached to the skull, and all signal processing was done by an external computer. Wires passing through the skin promote infection, so for a clinical device, the power into and neural signals out of the implant must be transmitted using telemetry across the skin. Unlike Hochberg and colleagues' prototype device, a clinical implant would therefore have to minimize data transfer by carrying out a considerable amount of signal processing within the implanted component of the neuroprosthetic. However, these types of problem are not insurmountable; indeed, they were overcome during the development of cochlear implants — the most successful neuroprosthetic so far (Box 1).

Animal-based studies tend to have three degrees of control or fewer because implantable neuromotor prosthetics are developed to predict an animal's movement when it reaches towards a spatial target. Human neuroprosthetics can use many more degrees of control because the level of involvement of the subject in calibrating the device will be so much greater — for instance, if the subject is asked to move each of their individual joints separately. As shown by Santhanam and colleagues, other control signals can be created based only on the patient's intentions but not actual actions. Importantly, neural processing in motor cortical regions is highly adaptable[10], probably reflecting our ability to acquire novel motor skills throughout life. So, using neuroprosthetic technologies should actually shape and modify brain processing to improve control of each prosthetic technology and facilitate the ability of paralysed individuals to convert their motor intentions into purposeful communication and interaction with the world. ■

Box 1 | Cochlear implants — the first successful neuroprosthetics

Human speech creates a complex spatio-temporal pattern of acoustic frequencies that is converted into neural activity by thousands of hair cells in the cochlea. Genetic factors, various diseases and pharmaceutical drugs can damage these cells, resulting in deafness (sensorineural hearing loss).

In the 1960s, implants began to be tested with the aim of using electrodes positioned within the cochlea to stimulate the neurons that are connected up to dysfunctional hair cells. Originally the devices included only a single stimulating electrode, but further advances led to four and then 22 electrodes being used by the 1980s. These devices have an external headpiece and processor that pick up sound from the environment and convert it into patterns of digital signals. These are transmitted through the skin to the implant attached to the skull and converted into a pattern of stimulation for each electrode.

It had been assumed that speech perception required much of the detailed pattern of acoustic frequencies to be sensed by the hair cells[11], suggesting that such a small number of independent sites of stimulation on the cochlea would provide only modest improvements in hearing — the initial devices were developed merely with the hope that they would aid lip reading. In practice, however, implant users were often surprisingly successful at understanding speech, forcing a re-evaluation of theories on speech perception.

More than 100,000 people now have cochlear implants, and many of these are young children who are given implants before they learn to speak to ensure the best development of auditory and speech skills. I can testify to the success of this neuroprosthetic technology and the improvement it makes in the quality of life of deaf individuals, as my son has a cochlear implant. He nonetheless attends classes in a regular school and has average to above-average auditory and speech skills compared with 'hearing' children his age.

It may be that other neuroprosthetic devices will surprise us by the extent of their usefulness, and provide unexpected insights into how the brain processes thoughts. **S.H.S.**

Stephen H. Scott is at the Centre for Neuroscience Studies, Department of Anatomy and Cell Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada.
e-mail: steve@biomed.queensu.ca

1. Hochberg, L. R. *et al. Nature* **442,** 164–171 (2006).
2. Santhanam, G., Ryu, S. I., Yu, B. M., Afshar, A. & Shenoy, K. V. *Nature* **442,** 195–198 (2006).
3. Wolpaw, J. R. & McFarland, D. J. *Proc. Natl Acad. Sci. USA* **101,** 17849–17854 (2004).
4. Masallam, S., Corneil, B. D., Gregor, B., Scherberger, H. & Andersen, R. A. *Science* **305,** 162–163 (2004).
5. Wessberg, J. *et al. Nature* **408,** 361–365 (2000).
6. Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R. & Donoghue, J. P. *Nature* **416,** 141–142 (2002).
7. Taylor, D. M., Tillery, S. I. & Schwartz, A. B. *Science* **296,** 1817–1818 (2002).
8. www.nature.com/nature/journal/v442/n7099/suppinfo/nature04970.html
9. Kennedy, P. R., Bakay, R. A., Moore, M. M., Adams, K. & Goldwaithe, J. *IEEE Trans. Rehabil. Eng.* **8,** 198–202 (2000).
10. Fetz, E. E. & Finocchio, D. V. *Exp. Brain Res.* **23,** 217–240 (1975).
11. Clark, G. M. *Cochlear Implants Int.* **1,** 1–17 (2000).

©2006 Nature Publishing Group

QUANTUM PHYSICS

New spin on the Hall effect

Andrew D. Kent

The spin Hall effect occurs when electrons with opposite spins go their separate ways in an electric field. The phenomenon is crucial to spin-based electronics, and its electrical signal has just been spotted.

Just as manipulating electron charge is central to conventional electronic circuitry, so creating, moving and measuring electron spin is essential to spin-based electronics, or spintronics[1]. On page 176 of this issue[2], Valenzuela and Tinkham report a breakthrough in spin control, with the first detection by electrical means of the 'spin Hall effect' — the separation, or polarization, of electrons of opposite spin (up and down) under the influence of an electric field.

Normally, electron spins are polarized by magnetic fields. Electrons in so-called ferromagnetic materials (iron is the classic example) are naturally spin-polarized; generating currents of spin in non-magnetic materials has therefore typically involved careful engineering of ferromagnetic contacts that can act as a source of polarized electrons. Alternatively, in semiconductors, polarized laser light has been used to create spin polarization.

But in fact, spin currents may appear wherever there are conventional charge currents[3–5]. This comes about through the interplay of an electron's spin and its direction of propagation, an effect known as the spin–orbit interaction. An electron moving forwards in an electric field experiences a magnetic field, with the result that the trajectories of spin-up and spin-down electrons are bent in different directions. A spin current and a spin imbalance are induced: this is the spin Hall effect.

So how can we detect this effect? In the conventional Hall effect, discovered in 1879 by Edwin Hall, positive and negative charges are separated by a transverse magnetic field, creating a voltage at right angles to the current flow. The spin Hall effect, however, does not naturally produce such an obvious electrical signal. Two research groups have detected the spin imbalance using optical methods[6,7]. But an electrical signature would open new opportunities for the study and control of spin currents.

Valenzuela and Tinkham's detection[2] relies on the fact that, if a current is initially unpolarized as it flows through a sample of material, the spin–orbit interaction causes equal numbers of electrons to end up on each side, and no transverse voltage results (Fig. 1a). But if the current is initially spin-polarized, different numbers of electrons migrate to either side of the sample. In this case, a charge imbalance results — and so a voltage. Thus, the key to the successful electrical detection of the spin Hall effect is creating a spin current in an otherwise non-magnetic material.

In the case in hand[2], that material is aluminium. The authors use a set-up known as a tunnel junction, consisting of a ferromagnetic electrode from which electrons, polarized in the direction of the electrode's magnetization, are injected into an aluminium strip through a thin intervening insulating layer. This injection of spin-polarized electrons creates a spin imbalance in the aluminium strip that must be counterbalanced by a flow of spins: a spin current is induced in an open circuit where no charge current is flowing[8,9] (Fig. 1b). In this region, voltage contacts are laid out in a 'Hall cross' geometry, enabling measurement of a transverse voltage.

In the authors' experimental geometry, the spins must have a component pointing at right angles to the plane of the strip for any effect to be observed. And indeed, when Valenzuela and Tinkham applied a magnetic field to force the magnetization of the ferromagnetic electrode into a perpendicular orientation, voltage contacts registered a transverse voltage that is the signature of the spin Hall effect.

The voltage signal decayed exponentially with the distance between the tunnel junction and the Hall cross. That is an expected consequence of spin–orbit coupling: as spins precess around an axis that varies with the electrons' direction of propagation, their initial polarization decays. Thus the spin current, and the resulting voltage, will diminish over a characteristic length, known as the spin diffusion length. Importantly, the authors find that the decay of the spin Hall voltage they measure is consistent with their independent measurements of the spin diffusion length.

The spin Hall voltage is small — of the order of 10 nV in this experiment, equivalent to a conductivity some 10^{-4} times smaller than values for conventional electrical conductivity. But it is easily measurable, as is the conventional Hall effect in metals. Soon after the discovery of that effect, an article in the popular press stated that "The new force is exceedingly feeble, so that we cannot predict any practical applications for it"[10]. That author was not to know of the effect's many uses in precision sensors for fluid flow, power and pressure in the era of integrated circuits, or in sensitive Hall magnetometers. The quantum-mechanical version of the Hall effect is now also used as the standard measure of resistance.

The spin Hall effect could be similarly useful. Now that electrical measurements of the effect are possible, we can expect to gain a deeper understanding of spin–orbit interactions in metals and semiconductors. Fundamental questions about the origin of the spin Hall effect that have generated considerable debate — such as whether the effect might be intrinsic, and so occur in pure materials, or whether it is always associated with impurities[11,12] — can also be addressed. Furthermore, transverse voltages much larger than those in the aluminium strips used by Valenzuela and Tinkham[2] should appear in semiconductor structures. Thus, the potential for generating and detecting spin currents electrically looks promising. ■

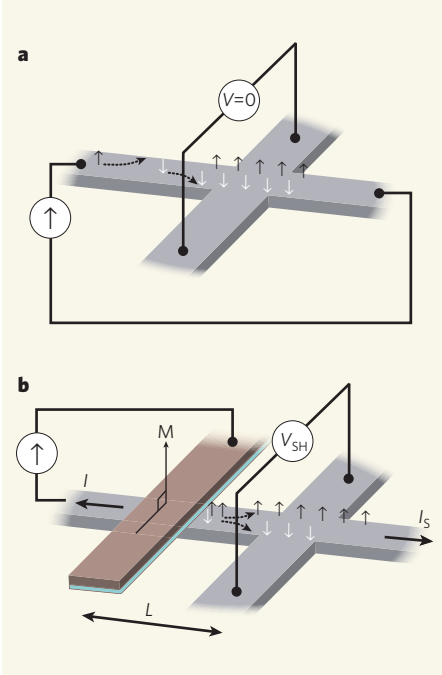


Figure 1 | Seeking the spin Hall effect. a, Charge current flows from left to right through a so-called Hall cross made from a conductor such as aluminium. Spin–orbit interactions will cause a separation of electron spins — the spin Hall effect. If the charge current is unpolarized (with equal numbers of spin-up and spin-down electrons), the spin imbalance does not induce a charge imbalance or transverse voltage at the Hall cross. **b**, If electrons polarized in the direction of magnetization M are injected from a ferromagnetic electrode while a circuit drives a charge current (I) to the left, a spin imbalance is created. This produces a spin current (I_S) without a charge current to the right of the electrode. Spin–orbit interactions again separate spin-up and spin-down electrons, but now the numerical superiority of one spin type means that a transverse charge imbalance and a spin Hall voltage, V_{SH}, are created. As the distance, L, between the electrode and the Hall cross increases, the voltage signal decreases, allowing the decay length of spin currents (spin diffusion length) to be measured.

Andrew D. Kent is in the Department of Physics, New York University, 4 Washington Place, New York, New York 10003, USA.
e-mail: andy.kent@nyu.edu

©2006 Nature Publishing Group

1. Wolf, S. A. *et al. Science* **294,** 1488–1495 (2001).
2. Valenzuela, S. O. & Tinkham, M. *Nature* **442,** 176–179 (2006).
3. Dyakonav, M. I. & Perel, V. I. *JETP Lett.* **13,** 467–469 (1971).
4. Hirsch, J. E. *Phys. Rev. Lett.* **83,** 1834–1837 (1999).
5. Zhang, S. *Phys. Rev. Lett.* **85,** 393–396 (2000).
6. Kato, Y. K., Myers, R. C., Gossard, A. C. & Awschalom, D. D. *Science* **306,** 1910–1913 (2004).
7. Wunderlich, J., Kaestner, B., Sinova, J. & Jungwirth, T. *Phys. Rev. Lett.* **94,** 047204 (2005).
8. Jedema, F. J., Heersche, H. B., Filip, A. T., Baselmans, J. J. A. & van Wees, B. J. *Nature* **416,** 713–716 (2002).
9. Johnson, M. & Silsbee, R. H. *Phys. Rev. Lett.* **55,** 1790–1793 (1985).
10. *The Nation* 44 (25 December 1879).
11. Murakami, S., Nagaosa, N. & Zhang, S.-C. *Science* **301,** 1348–1351 (2003).
12. Sinova, J. *et al. Phys. Rev. Lett.* **92,** 126603 (2004).

ATMOSPHERIC CHEMISTRY

Radicals follow the Sun

Paul O. Wennberg

Hydroxyl free radicals are part of a complex network of atmospheric chemical reactions. But a long-term study shows that their concentration can be predicted by the intensity of ultraviolet sunlight alone.

In an impressive feat of endurance and analytical skill, Rohrer and Berresheim have obtained a five-year record of the atmospheric concentration of the hydroxyl free radical, OH, at a research station in southern Germany (page 184 of this issue)[1]. This is remarkable because, although OH controls the rate of oxidation of many gases in Earth's atmosphere, the volume mixing ratio of this highly reactive (and thus short-lived) free radical is less than one part per trillion. Measuring its concentration, even over short periods, is a tremendous technical challenge.

The hydroxyl radical is central to Earth's atmospheric chemistry. For example, the large amounts of methane (CH_4) and carbon monoxide (CO) added to the atmosphere (0.6 billion and 2.5 billion tonnes per year, respectively) are offset by equally large sinks of these compounds from their reactions with OH (Fig. 1). In the absence of OH, the concentrations of these greenhouse gases would be more than an order of magnitude greater. The reactions of OH with carbon monoxide and with hydrocarbons are also responsible for producing most of the ozone (O_3) present in Earth's troposphere (the lowest part of Earth's atmosphere that extends 8–17 km from Earth's surface).

It was first suggested in 1971 that OH controls Earth's tropospheric chemistry[2]. Researchers immediately set out to measure tropospheric OH concentrations and the abundances of the various chemicals thought to affect OH levels. However, measuring OH concentration was much more challenging than expected, and it is only in the past few years that analytical methods of sufficient precision and accuracy have become available to do this[3]. These measurements have typically been made in 'campaign mode', where aircraft or ground-based sites are equipped with instruments for a short period of time. The data set described by Rohrer and Berresheim[1] adds a long-term view of atmospheric chemistry, and was collected at a well-equipped and established research site — the Hohenpeissenberg Meteorological Observatory, 1,000 m above sea level.

Analysis of these data[1] brings the authors to a striking conclusion — that variations in the concentration of OH may be explained, in a statistical sense, by a single parameter, even

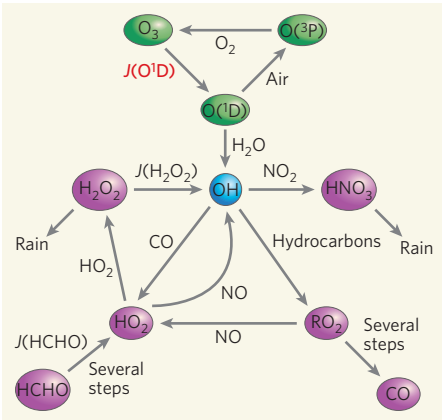


Figure 1 | The formation and chemistry of atmospheric hydroxyl radicals. The main route for hydroxyl-radical (OH) generation is shown in the green ovals. Ozone is broken up by ultraviolet light to form atomic oxygen in the ^1D excited electronic state; this occurs with a frequency denoted by $J(O^1D)$. The excited oxygen atoms react with water vapour to form OH radicals. Alternatively, the excited atoms may be recycled back to ozone, via the ^3P electronic ground state. The concentration of OH is usually less than 1 part per trillion per volume of gas. This reflects the rate of OH formation as described above; the rate of OH losses in reactions with, for example, nitrogen dioxide (NO_2), carbon monoxide (CO) and hydrocarbons such as methane; and the efficiency with which OH is recycled from its photochemical siblings HO_2 and RO_2 (where R is any hydrocarbon chain) through reactions with the common pollutant nitric oxide (NO). Recycling with NO is the mechanism responsible for the formation of ozone in the lower atmosphere. $J(H_2O_2)$ and J(HCHO) are the frequencies with which the respective molecules are fragmented by ultraviolet radiation.

though OH levels are expected to depend on a complex mixture of variables. The crucial parameter is a measure of the frequency with which a given molecule of ozone is broken up by solar ultraviolet radiation, yielding atomic oxygen in an excited electronic state — a parameter denoted as $J(O^1D)$ by atmospheric chemists, where ^1D is the excited state (Fig. 1). This frequency is determined by the intensity of ultraviolet radiation in a narrow spectral window between 305 and 330 nm (ref. 4) — the same light that causes sunburn. The solar intensity at these wavelengths is inversely related to the amounts of ozone, clouds and particulate matter between the Sun and the ground.

It is not surprising that the tropospheric concentration of OH depends on $J(O^1D)$. Many molecules are fragmented by ultraviolet radiation at these wavelengths, so this parameter can be thought of as a broad measure of photochemical activity. What is surprising is that the concentration of OH seems to depend solely on a factor that is not directly related to the local chemical and physical environment. Even more remarkably, the measured concentration of OH correlates substantially better with $J(O^1D)$ than it does with the predicted concentration calculated from a state-of-the-art model. The model includes as constraints not only $J(O^1D)$, but also the measured concentrations of most of the gases shown in Figure 1. This is disheartening, as it clearly illustrates that we still do not have an adequate description of tropospheric OH chemistry.

It is possible to read too much into the strong correlation between OH concentration and $J(O^1D)$. As Rohrer and Berresheim's analysis shows[1], this correlation probably results from correlated changes in other controlling factors, which balance each other out. For example, the atmospheric lifetime of OH is determined by the rate of the radical's destruction in its reactions with hydrocarbons (here, primarily biogenic compounds) and with certain pollutants. In the air at Hohenpeissenberg, the concentrations of biogenic compounds are higher in summer, whereas the concentrations of reactive pollutants, such as carbon monoxide and nitrogen dioxide (NO_2), are higher in winter. As a result, the seasonal variability in the atmospheric lifetime of OH at this site is very small. Similarly, the amount of OH produced from atomic oxygen in an excited state (O^1D) is high in summer because of the increased humidity, whereas the recycling of OH from related radicals is more efficient in winter because of the higher nitric oxide (NO) levels (Fig. 1).

The findings at Hohenpeissenberg are reminiscent of studies made in the lower stratosphere, the layer of Earth's atmosphere that sits above the troposphere. Here, OH concentrations were found to be highly dependent upon another single variable — the height of the Sun above the horizon[5]. Just as at Hohenpeissenberg, serendipitous correlations between the controlling factors of OH abundance reduced the apparent influence of these

©2006 Nature Publishing Group

factors[6]. Moreover, at the time of the measurements, the observed OH concentrations correlated better with the position of the Sun than with predicted concentrations calculated from the existing state-of-the-art photochemical model[7]. The inability to accurately calculate stratospheric OH levels spurred a re-examination of many of the pathways leading to OH production[8,9] — including the ultraviolet-induced destruction of ozone, described by $J(O^1D)$ (refs 4, 10). Descriptions of stratospheric chemistry were much improved by these studies, and eventually led to quantitative agreement between the observed and calculated OH concentrations[11].

The history of the stratospheric investigation provides a template for further research in the troposphere[12]. As illustrated by Rohrer and Berresheim[1], the chemical drivers of tropospheric OH production and loss are often correlated, masking their individual influence. Nevertheless, to define the global distribution of OH, its dependence on climate change, and the effect of industrial and agricultural emissions upon it, we must unmask the effects of these processes. Measurements such as those obtained by Rohrer and Berresheim will be essential for identifying errors and omissions in our theories of tropospheric oxidation chemistry. ■

Paul O. Wennberg is at the California Institute of Technology, Division of Geological and Planetary Sciences and Division of Engineering and Applied Science, 1200 East California Boulevard, Pasadena, California 91125, USA.
e-mail: wennberg@caltech.edu

1. Rohrer, F. & Berresheim, H. *Nature* **442,** 184–187 (2006).
2. Levy, H. *Science* **173,** 141–143 (1971).
3. Heard, D. E. & Pilling, M. J. *Chem. Rev.* **103,** 5163–5198 (2003).
4. Matsumi, Y. *et al. J. Geophys. Res.* **107,** 4024 (2002).
5. Wennberg, P. O. *et al. Science* **266,** 398–404 (1994).
6. Hanisco, T. F. *et al. J. Phys. Chem. A* **105,** 1543–1553 (2001).
7. Salawitch, R. J. *et al. Geophys. Res. Lett.* **21,** 2551–2554 (1994).
8. Hanson, D. R. & Ravishankara, A. R. *Geophys. Res. Lett.* **22,** 385–388 (1995).
9. Roehl, C. M., Nizkorodov, S. A., Zhang, H., Blake, G. A. & Wennberg, P. O. *J. Phys. Chem. A* **106,** 3766–3772 (2002).
10. Michelsen, H. A. *et al. Geophys. Res. Lett.* **21,** 2227–2230 (1994).
11. Salawitch, R. J., Wennberg, P. O., Toon, G. C., Sen, B. & 7Blavier, J. F. *Geophys. Res. Lett.* **29,** 1762 (2002).
12. Olson, J. R. *et al. J. Geophys. Res.* **111,** D10301 (2006).

PALAEONTOLOGY

A ghost with a bite

Stefan Bengtson

Witness a snail scraping microbial films from the inside of an aquarium. Go back 505 million years, and this looks to have been the way an enigmatic early animal made its living (but without the aquarium).

A faint shade on a piece of shale from British Columbia, Canada, has haunted palaeontology for 30 years. This is the fossil of *Odontogriphus*, the half-a-billion-year-old 'toothed riddle' from the Cambrian Burgess Shale, which has never really found peace within the evolutionary scheme of animals. With a poorly preserved body, it was mainly known through peculiarly arranged tooth-like spines, hypothesized to be the stiff supports of a cluster of tentacles. On page 159 of this issue, Caron *et al.*[1] present new fossils of the 'toothed riddle'. *Odontogriphus* at last comes into clear view, with a firm body and prominent teeth. This view suggests that it is allied to primitive molluscs, and so provides fresh insight into the early evolution of animals.

Odontogriphus was brought to the world's attention in 1976 by Simon Conway Morris[2]. The single specimen had been collected some 60 years earlier by Charles Walcott, the discoverer of that famous spot in the Burgess Pass where organisms from the dawn of animal life are found with soft tissues preserved in full flattened splendour. Even taking into account the flattening resulting from compression of the embedding sediment, *Odontogriphus* as reconstructed looked like a roadkill: a roughly parallel-sided body, about 4 cm long, blunt at both ends, with transverse stripes over most of the surface and with a body thickness only a fraction of its width.

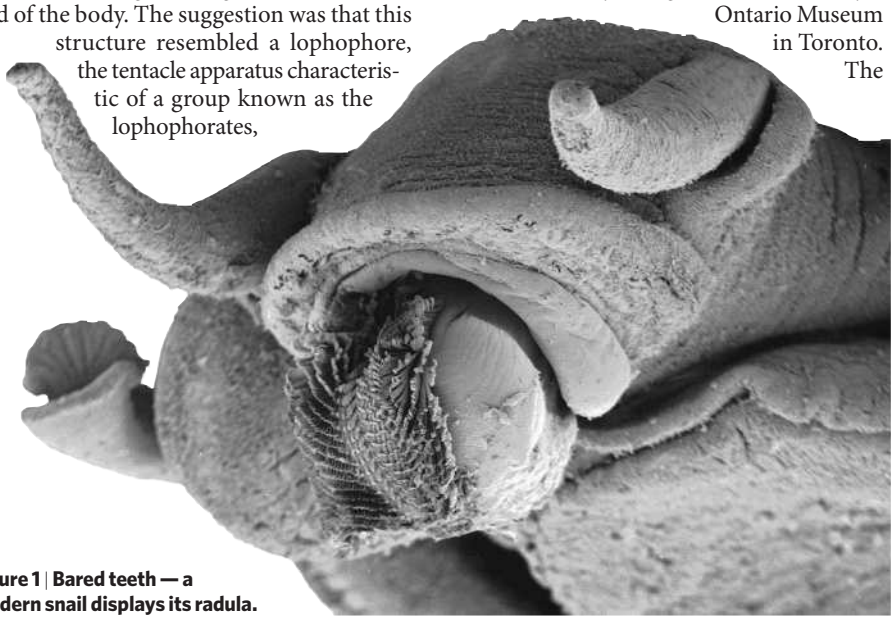
The distinctive feature, however, was a structure like a figure of eight near the blunter end of the body. The suggestion was that this structure resembled a lophophore, the tentacle apparatus characteristic of a group known as the lophophorates, which otherwise incorporate such different beasts as the bivalved brachiopods ('lampshells'), the tube-dwelling phoronids and the colonial bryozoans. Lophophorate animals feed by filtering water drawn into this apparatus by cilia.

Odontogriphus did not display any obvious tentacles, but, given the vagaries of fossilization, such a deficiency need not stop a palaeontologist. Where the tentacles were to have been expected, there were tooth-like structures, which could be interpreted as stiff supports for tentacles that had failed to be preserved. The conodont animal, a classical enigma of palaeontology, then entered the picture. The isolated, tooth-shaped fossils known as conodonts had been proposed to be not teeth, but the mineralized supports of tentacles. Could *Odontogriphus* provide the link by showing conodont-like 'teeth' arranged in the shape of a lophophore?

Fortunately for our navigation in this sea of red herrings, convincing fossils of conodont animals were soon thereafter identified[3], shredding any notions of resemblance to *Odontogriphus*. The latter still hovered around as a possible lophophorate, however, though more like a ghost than a real animal. There simply weren't enough data to make it materialize.

Other discoveries of wonderfully preserved Cambrian fossils then started to compete for attention[4–6], and by the early 1990s the zest seemed to have largely evaporated from studies of the Burgess Shale. Even the substantial collections made by Walcott early in the twentieth century and kept at the Smithsonian Institution in Washington DC had been pored over to such an extent that the law of diminishing returns started to take effect.

That tide was turned by palaeontologist Desmond Collins and his co-workers. Over several years, their tireless work in the mountains of British Columbia extracted new and wonderful material from the fabled shale, which they brought home to the Royal Ontario Museum in Toronto. The



A. WARÉN

Figure 1 | Bared teeth — a modern snail displays its radula.

©2006 Nature Publishing Group

new examples of *Odontogriphus* are among the fruits of that labour.

Caron *et al.*[1] have now had a good look into the mouth of *Odontogriphus*. They see a set of chevron-shaped bars, set with sharp teeth. Each animal has at least two chevrons, which seem to stick tightly together even after death and decomposition. Sometimes the foremost chevron has flipped over, the two then producing a pinched shape vaguely like that figure of eight. In addition, however, many specimens show additional chevrons, faintly expressed, behind the two distinct front ones. This, to Caron *et al.*, is a giveaway for a radula, the tooth apparatus found in many living molluscs (Fig. 1). Witness a snail scraping microbial films from the inside of an aquarium glass. It does so with teeth positioned in rows on a ribbon, and as the front teeth wear out they get thrown out (or, not uncommonly, swallowed) and the ribbon advances to expose fresh sets of teeth.

Other features of *Odontogriphus* also point towards a mollusc affinity. A structure like a muscular sole lies on what was presumably the lower side, and the transverse stripes observed on the first specimen now seem to be wrinkles on that sole. Furthermore, the sole is surrounded by serially repeated dark structures that Caron *et al.* interpret as serial gills similar to the ones that occupy the crease between body and sole in chitons, the eight-plated molluscs we see clinging to rocks on tidal shores.

The great match, however, is not so much with known and accepted molluscs as with *Wiwaxia*, another Burgess Shale fossil of even greater repute than *Odontogriphus*. *Wiwaxia* is a scale-clad conundrum that in evolutionary reconstructions usually moves around among the early lineages of the lophotrochozoans (a group incorporating lophophorates, molluscs and annelids)[7,8], but it has even been interpreted as an annelid of modern type[9]. Its scaly armour resembles that shared by a disparate group of Cambrian animals known as coeloscleritophorans[10], but its mouth apparatus is nearly indistinguishable from that of *Odontogriphus*.

Is there a lesson to be learned from this mosaic of resemblances? In terms of evolutionary time, Cambrian animals are all closely related to one another, and the first place to look for similarities should be in other organisms from the same time. Comparisons with animals living 500 million years later, although crucial for our understanding of how the living world came to be, cannot pick up the lost treasures of the fossil record, be they lineages, organisms or characters.

A final kink to the new story is that Caron *et al.*[1] point to similarities between *Odontogriphus* and an even older mollusc-like organism, *Kimberella*[11], from the late Precambrian about 555 million years ago. *Kimberella* doesn't show any teeth, but the fossil is commonly associated with sweeping marks on the surrounding microbial mats[12], much like those produced by that snail on the inside of the aquarium glass. The bite of the ghost may be deep indeed. ■

Stefan Bengtson is in the Department of Palaeozoology, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden.
e-mail: stefan.bengtson@nrm.se

1. Caron, J.-B., Scheltema, A., Schander, C. & Rudkin, D. *Nature* **442,** 159–163 (2006).
2. Conway Morris, S. *Palaeontology* **19,** 199–222 (1976).
3. Briggs, D. E. G., Clarkson, E. N. K. & Aldridge, R. J. *Lethaia* **16,** 1–14 (1983).
4. Müller, K. J. & Walossek, D. *Trans. R. Soc. Edinb. Earth Sci.* **76,** 161–172 (1985).
5. Conway Morris, S., Peel, J. S., Higgins, A. K., Soper, N. J. & Davis, N. C. *Science* **326,** 181–183 (1987).
6. Hou, X. & Sun, W. *Acta Palaeontol. Sinica* **27,** 1–12 (1988).
7. Conway Morris, S. *Phil. Trans. R. Soc. Lond. B* **307,** 507–586 (1985).
8. Eibye-Jacobsen, D. *Lethaia* **37,** 317–335 (2004).
9. Butterfield, N. J. *Paleobiology* **16,** 287–303 (1990).
10. Bengtson, S. in *Evolving Form and Function: Fossils and Development* (ed. Briggs, D. E. G.) 101–124 (Yale Peabody Mus., New Haven, CT, 2005).
11. Fedonkin, M. A. & Waggoner, B. M. *Nature* **388,** 868–871 (1997).
12. Seilacher, A. *Palaios* **14,** 86–93 (1999).

SOLID-STATE PHYSICS

Supersolid simulations

Dieter Jaksch

Supersolids — substances that are crystalline but also behave as free-flowing superfluids — can exist, according to quantum theory. Models now suggest a route to the clinching experimental evidence.

Quantum theory predicts the existence of several phases of matter that have counter-intuitive properties. Some of these phases — such as superconductors, through which electricity flows without resistance, and superfluids, which flow without friction — have been experimentally verified and have assumed major roles in science and technology. Others, such as atomic supersolids, are more elusive. Supersolids have seemingly mutually exclusive properties: they are rigid crystals with well-defined, long-range spatial order, but they can also behave as if they are superfluids.

Writing in *Physical Review A*, Scarola and colleagues[1] propose a method for detecting supersolids by analysing the quantum interference that occurs between atoms when released from a web of confining laser light known as an optical lattice. The great advantage of this technique is that it measures the defining signatures of superfluidity and spatial order directly, and could therefore answer long-standing questions on the existence and properties of supersolid matter.

The history of possible physical mechanisms leading to supersolidity can be traced to 1969, when Andreev and Lifshitz[2] investigated the role of 'vacancy sites', at which an atom is absent in a crystal. They proposed that vacancies could delocalize over the crystal by forming a Bose–Einstein condensate (a shared quantum state) at very low temperatures. This phenomenon could give rise to superfluid currents that would not spoil the regular ordering of the atoms.

This effect was later shown also to be present in the phase diagram of so-called extended-lattice Hubbard models at a temperature of absolute zero. The phases in these models occur through competition between particle tunnelling between neighbouring lattice sites, which acts to delocalize particles, and interactions between particles, which favour localized states. Two different types of interaction are present in extended Hubbard models: 'on-site', between particles occupying the same lattice site as a result of tunnelling; and 'off-site', in which particles are affected by others in nearby sites.

Off-site interactions are crucial for the existence of a supersolid phase. It is relatively simple to write down quantum-mechanical wavefunctions of Hubbard models that correspond to substances displaying supersolidity. But determining whether these states can actually be attained experimentally in macroscopic crystalline samples, and whether the states are thermodynamically stable, is much harder.

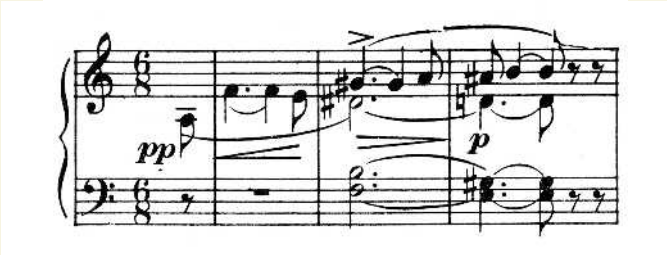
Indeed, attempts to observe supersolidity in the laboratory, mostly using the helium isotope ^{4}He, were unsuccessful for 35 years until the first experimental evidence[3] was reported in 2004. In that experiment, a solid ^{4}He sample was rotated to and fro in a torsional oscillator. When the sample was cooled to below 175 millikelvin, a decrease in its rotational inertia was detected, indicating that about 1% of the sample stopped oscillating with the solid. The origin of this superfluid component, in particular whether it is truly a state of matter in thermodynamic equilibrium — and therefore long-lasting — is still unclear. The dependence of superfluidity on added ^{3}He and ambient pressure, as well as the fact that the superfluid component is reduced by annealing (warming and then slowly re-cooling the sample), are also not well understood, and require further experimental and theoretical insights.

So what can optical lattices contribute to the physics of supersolids? Such lattices are created by the interference of laser beams

©2006 Nature Publishing Group

MUSIC

Calculated tones



The opening chord of Richard Wagner's *Tristan und Isolde* (in the third bar of the score pictured) caused quite a stir when the opera was premièred in 1865. The daring combination of tone intervals — the dissonant augmented fourth; the dynamic, imperfectly consonant major third; the highly consonant perfect fourth — came to be seen as a key change in attitudes to harmony, an overture to the atonality of much subsequent classical music.

The question of how simultaneous notes combine to form aesthetically pleasing (or displeasing) harmonies, and how each harmony best evolves to the next one — a technique known as counterpoint, or voice leading — has exercised music theorists for centuries.

Dmitri Tymoczko (*Science* **313,** 72–74; 2006) is the latest to attempt to systematize such relationships mathematically, with the goal of establishing a universal 'geometry' of musical chords.

A musical octave separates one note from another similar-sounding note of double (or half) the frequency. In the Western tradition, each octave is divided into 12 distinctive tones. Tymoczko therefore uses a simple logarithmic expression to convert the fundamental frequency of each note to a real number in 'pitch space', starting from C (which is 0) and Csharp (1) and going to A (9), Bflat (designated as *t*) and B (*e*). After B comes C, which, as the start of the next octave, is 0 again.

The geometrical distance in pitch space between two chords can then be assessed through a 'voice-leading size', a relative measure of how much the constituent notes in one chord must change to make the second. The results can be used to form a geometrical space, called an orbifold, on which the separation between any two chords is proportional to their voice-leading size.

Because of the cyclical nature of the octave, the orbifold loops back on itself: for two-tone chords, for example, it resembles a Möbius strip — with a half-twist along its circumference — whose boundaries, for the purposes of voice leading, act like a mirror. An equivalent picture applies in higher dimensions (corresponding to chords with more notes), creating a unified framework for considering all possible chord combinations.

So what does this teach us? It seems that many of the chord progressions favoured by composers in constructing pleasing counterpoints exploit symmetries of the orbifold's geometry. Chords that divide an octave almost evenly are said to be near transpositional symmetry, and include the familiar consonant chords. These chords cluster into the centre of the orbifold. Nearly permutationally symmetric chords, on the other hand, which involve notes that are close together and therefore sound dissonant, collect towards the boundaries.

A third class of chord, known as nearly inversionally symmetric, is scattered throughout the orbifold. Such chords are prominent in both tonal and atonal music — and particularly in the works of such nineteenth-century pioneers as Richard Wagner.

Richard Webb

that cross at an angle, setting up a regular intensity pattern of light in space. This pattern acts as a periodic potential that can trap cold, neutral atoms. The resulting 'optical crystal' is virtually defect-free and, by loading atoms into it from a Bose–Einstein condensate, it can be used to achieve very low temperatures and well-defined densities.

The versatility of the lattice laser arrangement and the controllability of the on-site atomic interactions provide a 'tool-box'[4] of Hubbard models for different systems. These models have well-known ranges of validity and their derivation is based on detailed understanding of the underlying microscopic dynamics of each system. This tool-box has already been used to investigate the transition of a superfluid to a Mott insulator (a low-temperature insulating phase in which the atoms occupy well-localized states), and the Tonks–Girardeau gases that under certain conditions emerge from Bose–Einstein condensates[5].

Even more exotic states predicted by extended-lattice Hubbard models might soon be realized if strong, adjustable off-site interactions can be obtained by loading molecules composed of atoms of different chemical elements, instead of single atoms, into a lattice[6]. Given that a supersolid phase is present in the model phase diagram, it might therefore be possible to obtain an unambiguous experimental demonstration of a supersolid in an optical crystal.

Scarola *et al.*[1] tackle the question of how to detect the supersolid phase in an optical lattice. They suggest using an existing experimental method[7] in which ultracold atoms are suddenly released from a lattice. The resulting interference pattern between the freely expanding quantum waves of the released atoms would display an additional peak if a supersolid component displaying superfluidity were present in the lattice. Simultaneously, statistical variations in the atom density (known as shot-to-shot noise) would provide information on the spatial ordering in the lattice, and so prove that the lattice system was a solid.

In practice, the situation is complicated by a confining potential, or 'trap', superimposed on the lattice that restricts the atoms to a certain region of space. This trap is usually harmonic (having a parabolic potential-energy form) and so produces fuzzy system boundaries that induce sizeable edge effects in the atomic interference pattern. Scarola *et al.* use a mean-field approximation that is known to capture the main features of the physics, and show that a two-dimensional supersolid can be detected even if a harmonic trap is present. They find that edge effects cause an additional interference peak similar to the one originating from the supersolid. However, the quantum-mechanical functions that characterize supersolid correlations between atoms are still quantitatively different from those in other phases. These quantitative features can be accurately accounted for within the model, and allow supersolidity to be indicated directly and unambiguously.

Quite apart from adding to our knowledge of ultracold atoms and the ever more exotic quantum phases they produce, such model experiments can also contribute to a better understanding of more complex systems. Puzzling properties of supersolid ^{4}He, such as its apparent dependence on the presence of ^{3}He, could be simulated by adding a second atomic species. Controlled addition of imperfections to the lattice and the mimicking of annealing processes might yield further insights into the phenomenon. Richard Feynman's vision[8] that quantum systems with exceedingly many degrees of freedom may be simulated by simpler, well-understood set-ups seems to be coming true. ■

Dieter Jaksch is in the Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.
e-mail: d.jaksch1@physics.ox.ac.uk

1. Scarola, V. W., Demler, E. & Das Sarma, S. *Phys. Rev. A* **73,** 051601 (2006).
2. Andreev, A. M. & Lifshitz, I. M. *Sov. Phys. JETP* **29,** 1107–1113 (1969).
3. Kim, E. & Chan, H. W. *Nature* **427,** 225–227 (2004).
4. Jaksch, D. & Zoller, P. *Ann. Phys.* **315,** 52–79 (2005).
5. Bloch, I. *Nature Phys.* **1,** 23–30 (2005).
6. Micheli, A., Brennen, G. K. & Zoller, P. *Nature Phys.* **2,** 341–347 (2006).
7. Fölling, S. *et al. Nature* **434,** 481–484 (2005).
8. Feynman, R. P. *Int. J. Theor. Phys.* **21,** 467–488 (1982).

©2006 Nature Publishing Group

NATURE|Vol 442|13 July 2006

OBITUARY

Raymond Davis Jr (1914–2006)

Father of solar neutrino detection.

For 30 years, Ray Davis could have been mistaken as a miner. Clad in hard hat, headlamp and battery belt, he would join 50 other 'first shifters' for a pre-dawn ride a mile into the Earth. A labyrinth of hot, dimly lit tunnels greeted the riders, but Davis knew his way, having retraced it some 700 times. Turning to the right, he headed towards the tunnel with the bright lights. At the end, beneath those lights, was a 380,000-litre tank. This was where he did his mining. While his companions blasted tonnes of granite in search of gold, Davis sifted through 600 tonnes of cleaning fluid for a dozen special atoms — much more interesting to him than gold.

His was the more difficult task. No mining engineer would accept the odds — find 10 atoms in a sea of 10^{31} — nor would workers a third his age long want to keep pace. He spent 15 hours a day underground each time he came to Homestake and, consequently, often wouldn't see the Sun. However, if you asked him about it, he would laugh, flash a broad smile and say, "I'm watching the Sun all the time!" Ray Davis, who died on 31 May, had indeed been watching the Sun in a new and interesting way since 1967. By doing so, he helped discover something unexpected in particle physics.

It was in the late 1930s that Hans Bethe provided the first rigorous explanation of why the Sun shines. Nuclear reactions in the core fuse hydrogen into heavier elements, with light being the most obvious (but not the only) by-product. Another particle, the neutrino, was also produced and, unlike light, travelled virtually unhindered from the solar interior. Wolfgang Pauli postulated this particle in 1930 to preserve several sacred laws of physics, but in the same breath dismissed it as probably undetectable given its tiny cross-section and resulting elusive nature. His worry was unfounded. In the 1950s, Fred Reines and Clyde Cowan saw signs of this poltergeist at the Savannah River reactor, an intense source of (anti)neutrinos.

Davis also began his neutrino work at reactors. He had earned a PhD at Yale, served in the army during the war, and worked briefly as a radiochemist at Monsanto, before joining the newly created Brookhaven National Laboratory in 1948 — his home for the next 37 years. When he asked the chemistry department chairman for his assignment, he was told to go to the library and find something interesting to work on. He did. A review paper on neutrinos caught his eye, as did a technical brief by Bruno Pontecorvo. Pontecorvo proposed a way

BROOKHAVEN NATL LAB.

of detecting neutrinos using a chlorine-containing liquid. The idea was that a neutrino would be captured on a particular chlorine isotope (^{37}Cl), changing it into a radioactive argon atom (^{37}Ar), which could be removed from the liquid and counted. Davis fleshed out the scheme, developing the extraction, gas handling and counting hardware needed to make it work. He constructed a 3,800-litre detector and looked for argon production at both the Brookhaven and Savannah River reactors. He found none — not because of any experimental failings, but because of his emerging finding that neutrinos and antineutrinos were different — at least with respect to capture on ^{37}Cl. In 1960, he began thinking about a grander challenge — detecting neutrinos from the Sun.

In this quest, Davis found an able guide in John Bahcall. Bahcall, a theorist, began calculating neutrino emissions from the Sun and their capture rate on chlorine. Davis was proposing a scaled-up 380,000-litre detector filled with C_2Cl_4 (a dry-cleaning fluid). Bahcall predicted that, of the more than 10^{22} neutrinos transiting this detector each week, about 10 would be captured, creating 10 argon atoms a week. Though small, it was a rate that Davis thought he could measure. Not one to make an idle boast, he had been working to improve the technologies for this large-scale detector and to understand its background. From earlier measurements, he knew that his detector would have to be deep underground to shield it from cosmic rays. After a brief search for a site, the detector was built at the Homestake Gold Mine in South Dakota from 1965 to 1967.

A problem arose almost immediately when data started arriving. Davis's measurements were about a third of Bahcall's prediction. In spite of more refined calculations and more data, the discrepancy remained. Dubbed the 'solar neutrino problem', it persisted for 20 years. Davis and his experiment remained in the crosshairs the entire time. He was challenged repeatedly — about efficiencies, chemical traps, pump changes, and the like — but he faced all questioners with his trademark candour, graciousness and wit. He brought technical precision and thoroughness to his answers, recounting the many careful, quantitative tests he had performed to address each concern. He had deep humility, but also quiet confidence in his experimental skills — which were formidable.

Beginning in the late 1980s, other experiments came online. They confirmed the problem. In the end, a versatile detector in Canada resolved it, showing that neither Bahcall's calculation nor Davis's mining was mistaken. Instead, neutrinos from the Sun were changing 'flavour' en route to Earth. This unexpected result meant that neutrinos, formerly thought to be massless, necessarily had some — a finding that required revision of the standard model of particle physics and gave these mysterious messengers a mass on a par with all visible matter in the Universe.

In 2002, Davis won the Nobel Prize in Physics. His wife of 54 years, Anna, and 21 other Davises accompanied him to Stockholm. Then 88 and battling Alzheimer's, Ray received an additional honour. Nobel protocol ordinarily requires each recipient to walk to the centre of the stage to receive their prize. However, on this occasion the King of Sweden came in person to a laureate. The gold that Davis had never sought, never mined for, found him that day.

Those of us who worked alongside Ray Davis, at Brookhaven and later at the University of Pennsylvania, saw up-close a great experimentalist — heroically persistent, patiently tenacious and meticulously precise. But we saw an even greater human being — one whose kindness, warmth and modesty were even more remarkable.

James R. Distel

James R. Distel is at Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.
e-mail: jdistel@lanl.gov

©2006 Nature Publishing Group

BRIEF COMMUNICATIONS

An atom-sorting machine

Laser-trapped atoms in strings can be deftly rearranged and the spacing between them precisely adjusted.

Laser cooling and trapping techniques allow us to control and manipulate neutral atoms[1]. Here we rearrange, with submicrometre precision, the positions and ordering of laser-trapped atoms within strings by manipulating individual atoms with optical tweezers[2]. Strings of equidistant atoms created in this way could serve as a scalable memory for quantum information[3].

Atoms and molecules on surfaces have previously been rearranged at the nanometre-scale using scanning tunnelling microscopy[4,5], and the first steps have been taken towards rearranging strings of trapped ions[6]. To realize a related technique in the context of cold-atom physics, we first trapped and cooled caesium atoms to sub-millikelvin temperatures inside a vacuum chamber using a magneto-optical trap[1]. These atoms were then transferred into a horizontal standing-wave optical dipole trap (HDT)[1], formed by two counterpropagating laser beams.

The dipole-trap lasers are far detuned from all caesium transitions. They polarize the atoms and pull them into the intensity maxima, located at the antinodes and separated by about 0.5 μm. The trapped atoms (storage time of up to 1 min) are then randomly distributed over the antinodes, forming a string along the axis of the trap. We determined the position of each atom from a fluorescence image, recorded using an intensified charge-coupled device (ICCD) camera[7,8].

By varying the relative phase of the trapping beams, the HDT can be moved along the x direction (Fig. 1a), shifting the string of atoms but leaving the interatomic separations unchanged[8,9]. To control these separations, we used a second, vertical standing-wave dipole trap (VDT) produced by retro-reflecting a focused laser beam (Fig. 1a).

In both dipole traps, the confining forces along the trap axis are much larger than in the radial direction. Accordingly, an atom stored in the overlap region of both traps will always follow the axial motion of the traps. This allowed us selectively to extract an atom from the HDT: after transporting it along the x direction into the overlap region of both traps, the atom is shifted upwards by axially moving the standing-wave pattern of the VDT (step 1 in Fig. 1a). The string of atoms left in the HDT can be moved to any position along the HDT with respect to the VDT (step 2, Fig. 1a). By reinserting the extracted atom into the HDT, it can then be placed at any target distance d_t from the remaining atoms (step 3, Fig. 1a).

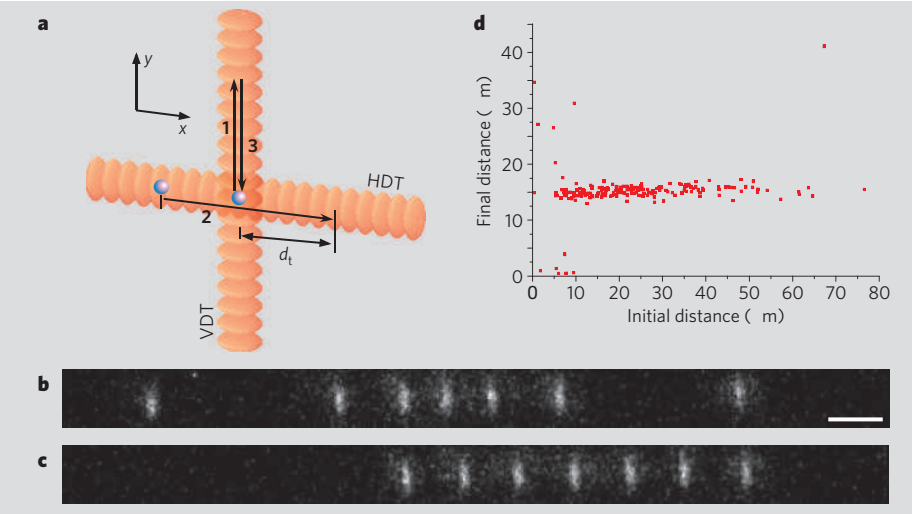


Figure 1 | Sorting atoms using optical tweezers. a, Two crossed standing-wave dipole traps are used to rearrange strings of trapped neutral atoms. (For methods, see supplementary information.) **b,** Fluorescence image of an initial string of seven randomly separated atoms. **c,** The same atoms after rearranging six of them using the distance-control operation. Scale bar, 15 μm. **d,** Relation between the final and the initial distance between simultaneously trapped atoms recorded for about 190 atom pairs. For initial distances greater than 10 μm, our distance-control operation successfully rearranges the atoms, resulting in a narrow distribution of final distances around the target distance $d_t = 15$ μm.

Through sequential application of this distance-control operation (duration, 400 ms; success rate, 98(−5/+2)%; see below), atoms can be arranged in equidistant strings. A detailed sequence of ICCD images illustrates this procedure for three atoms (see supplementary information); Fig. 1b, c shows ICCD images of a string of seven atoms before and after rearrangement, respectively. The initially random separations have been equalized ($d_t = 15$ μm).

The results shown in Fig. 1d characterize the performance of the distance-control operation: for initial distances of more than 10 μm, we obtain a final distance of 15.27 μm, with a standard deviation of 0.78 ± 0.05 μm. If the initial separation between atoms is less than 10 μm, extracting one of them leads to uncontrolled effects on the other atom (extraction, ejection and so on) owing to the 10-μm radius of the VDT.

Laser-trapped atoms are efficiently isolated from the environment and show little decoherence[3,10]. The equidistant strings of atoms produced by our method could therefore serve as a scalable, neutral-atom quantum register for storing and manipulating quantum information in the context of cavity quantum electrodynamics[11] or controlled cold collisions[12,13].

Yevhen Miroshnychenko, Wolfgang Alt, Igor Dotsenko, Leonid Förster, Mkrtych Khudaverdyan, Dieter Meschede, Dominik Schrader, Arno Rauschenbeutel
Institut für Angewandte Physik, Universität Bonn, 53115 Bonn, Germany
e-mail: arno.rauschenbeutel@iap.uni-bonn.de

1. Metcalf, H. J. & van der Straten, P. *Laser Cooling and Trapping* (Springer, New York, 1999).
2. Grier, D. G. *Nature* **424,** 810–816 (2003).
3. Schrader, D. *et al. Phys. Rev. Lett.* **93,** 150501 (2004).
4. Eigler, D. M. & Schweizer, E. K. *Nature* **344,** 524–526 (1990).
5. Bartels, L., Meyer, G. & Rieder, K.-H. *Phys. Rev. Lett.* **79,** 697–700 (1997).
6. Hensinger, W. K. *et al. Appl. Phys. Lett.* **88,** 034101 (2006).
7. Dotsenko, I. *et al. Phys. Rev. Lett.* **95,** 033002 (2005).
8. Miroshnychenko, Y. *et al. Opt. Expr.* **11,** 3498–3502 (2003).
9. Kuhr, S. *et al. Science* **293,** 278–280 (2001).
10. Kuhr, S. *et al. Phys. Rev. Lett.* **91,** 213002 (2003).
11. You, L., Yi, X. X. & Su, X. H. *Phys. Rev. A* **67,** 032308 (2003).
12. Jaksch, D., Briegel, H.-J., Cirac, J. I., Gardiner, C. W. & Zoller, P. *Phys. Rev. Lett.* **82,** 1975–1978 (1999).
13. Mandel, O. *et al. Nature* **425,** 937–940 (2003).

Supplementary information accompanies this communication on *Nature*'s website.
Received 23 February; accepted 11 May 2006.
Competing financial interests: declared none.
doi:10.1038/442151a

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see *Nature* contents.

©2006 Nature Publishing Group

EMBRYOLOGY

Does prepatterning occur in the mouse egg?

Arising from : B. Plusa *et al. Nature* 434, 391–395 (2005)

A recurring question in developmental biology has been whether localized determinants play any role in mammalian preimplantation development[1]. This is a controversial issue that brings back the idea of prepatterning[2–4] and is explored further by Plusa *et al.*[5], who claim it is the first cleavage of the mouse zygote that predicts the blastocyst axis, rather than the animal pole or sperm entry point, as previously suggested[3]. However, other evidence indicates that the blasotcyst axis is not predetermined and there is no prepatterning in the mouse egg[6–12]. Here we investigate the origin of these different views and conclude that they arise from differences in the data themselves and in their interpretation.

According to the prepatterning model[4], the first cleavage plane (defined by the second polar body and the sperm entry point[3]) manifests in the blastocyst as the boundary between the embryonic and abembryonic parts. The two-cell blastomere that inherits the sperm entry point tends to divide first to produce cells that populate predominantly the embryonic part of the blastocyst, and there is a stereotyped cleavage pattern up to the four-cell stage[13].

The features of this model have been challenged[6–12], however, reinforcing the view that the mammalian embryo is highly regulative and that prepatterning is absent in the mouse egg: there is no stereotyped cleavage pattern up to the four-cell stage[6,9,10]; in most blastocysts, the clones derived from the two-cell blastomeres do not correspond to the embryonic–abembryonic territories[7,8,10–12]; and the blastomere that inherits the sperm entry point is not different from the other[10,12]. We conclude that clonal separation during cleavage has no causal relationship to fate specifications.

The four-dimensional studies of Plusa *et al.*[5] give opposite results to ours[6]. However, we question the accuracy of *z*-axis measurement using a couple of scans, based on our experience of using 40–80 sections for a single three-dimensional analysis in a separate experiment[10]. Another concern is the condition of the embryos during the recordings: if the authors consistently used four-dimensional analysis with fluorescent visualization, our results indicate that the embryos would have been severely damaged, showing abnormal behaviour and compromised development, as a result of their exposure to high-energy light[14]. In our experiments, we kept light exposure to a minimum and verified that the recorded embryos developed to term[6,10].

We did not select the embryos[5], but aligned them using a manipulator (see Figs 2–4, but not Fig. 1, of ref. 6) so that the second polar body and the two pronuclei were all in the focal plane. In fact, our model of the first cleavage specification[6] is confirmed[1] in Fig. 3 of Plusa *et al.*[5]. Cytochalasin treatment was necessary for the pronuclear transfer[6], which represents only a small part of the study and confirms most of the data from normal embryos. Furthermore, the conditions used by Plusa *et al.*[5] differ from ours[6] in length of exposure (4 hours, compared with 10–15 min) and temporal proximity to cleavage (6 hours, compared with 10–12 hours).

The main conclusion of Plusa *et al.*[5] is based on their Fig. 4. This might be important if their analyses are unbiased with respect to their counting method[7,8,10,11]. A possible flaw in their counting method[5] is that the "boundary zone" is too large, being composed of one-third of the total cell number[7], and plays a buffering role: we cannot see the justification for choosing the number of three or fewer as evidence for predisposition.

The regulative model (Fig. 1) applies to any embryos, regardless of the conditions (cloning or chimaera aggregation, for example) or strains. By contrast, the prepatterning model explains polarity establishment (a fundamental process) by tendencies, and possible differences depending on strains and conditions. It also applies to only a small proportion of embryos (compared with the random hypothesis). The idea of embryonic–abembryonic territories[2–5] is biologically irrelevant as both contain inner-cell mass and trophectoderm. Segregation of the two lineages starting, but not yet determined, at the morula stage[15] is the essential process, enabling fluid accumulation and blastocoel formation. Thus, blastocyst morphogenesis is controlled by regulative cellular interactions integrating external mechanical cues (such as the zona pellucida shape[7,9,10]), if available, but not by intrinsic bias as in the prepatterning model[4], leading to the specification of the embryonic–abembryonic axis and four distinct lineages[8,10,11].

Preimplantation development leading to the formation of extraembryonic structures is unique to mammals. Patterning driven by differentially localized molecules, reminiscent of non-mammals, only starts in blastocysts. The regulative model provides a conceptual basis for understanding early mammalian development, and for clinical applications, including preimplantation genetic diagnosis.

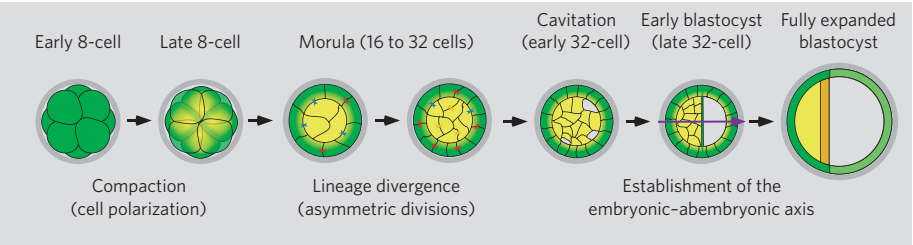


Figure 1 | Blastocyst development in mice. The first asymmetries related to the divergence of cell lineages appear at the eight-cell stage, when blastomeres polarize. Polar blastomeres divide symmetrically (daughter cells are polar; red link) or asymmetrically (one is polar and the other is non-polar; blue link) to generate an internal mass of non-polar cells that divide equally (orange link) to give rise to most pluriblast cells and an outer layer of polar cells that give rise mainly to trophoblast cells and some pluriblast cells (after asymmetric division). These asymmetric divisions allow the embryo to reduce the variability in the number of inner cells at the 16-cell stage. By the 32-cell stage, small cavities at the basal part of trophoblast cells coalesce to form the blastocoel, in which pluriblast cells (yellow), are separated from the cavity by trophoblastic processes (dark green), when the first embryonic axis (purple arrow) is set up. In fully expanded blastocysts, the trophoblastic processes regress and a layer of pluriblast cells facing the blastocoel differentiate into hypoblast (or primitive endoderm; orange). The polar trophoblast (dark green) is in contact with the epiblast (yellow) and the mural trophoblast (light green) surrounds the blastocoel.

Takashi Hiiragi*, Sophie Louvet-Vallée†, Davor Solter*, Bernard Maro†‡
*Max Planck Institute of Immunobiology, 79108 Freiburg, Germany
e-mail: hiiragi@immunbio.mpg.de
†Laboratoire de Biologie Cellulaire du Développement, UMR 7622, CNRS, Université Pierre et Marie Curie, 75252 Paris cedex 05, France
‡Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel

1. Rossant, J. & Tam, P. P. *Dev. Cell* **7,** 155–164 (2004).
2. Gardner, R. L. *Development* **128,** 839–847 (2001).
3. Piotrowska, K. & Zernicka-Goetz, M. *Nature* **409,** 517–521 (2001).
4. Zernicka-Goetz, M. *Development* **129,** 815–829 (2002).
5. Plusa, B. *et al. Nature* **434,** 391–395 (2005).
6. Hiiragi, T. & Solter, D. *Nature* **430,** 360–364 (2004).
7. Alarcon, V. B. & Marikawa, Y. *Biol. Reprod.* **69,** 1208–1212 (2003).
8. Chroscicka, A., Komorowski, S. & Maleszewski, M. *Mol. Reprod. Dev.* **68,** 308–312 (2004).
9. Louvet-Vallée, S., Vinot, S. & Maro, B. *Curr. Biol.* **15,**

© 2006 Nature Publishing Group

464–469 (2005).
10. Motosugi, N., Bauer, T., Polanski, P., Solter, D. & Hiiragi, T. *Genes Dev.* **19,** 1081–1092 (2005).
11. Alarcon, V. B. & Marikawa, Y. *Mol. Reprod. Dev.* **72,** 354–361 (2005).
12. Fujimori, T., Kurotaki, Y., Miyazaki, J. & Nabeshima, Y. *Development* **130,** 5113–5122 (2003).
13. Piotrowska-Nitsche, K. & Zernicka-Goetz, M. *Mech. Dev.* **122,** 487–500 (2005).
14. Squirrell, J. M., Wokosin, D. L., White, J. G. & Bavister, B. D. *Nature Biotechnol.* **17,** 763–767 (1999).
15. Vinot, S. *et al. Dev Biol.* **282,** 307–319 (2005).

doi:10.1038/nature04907

EMBRYOLOGY

Plusa et al. *reply*

Replying to: T. Hiiragi, S. Louvet-Vallée, D. Solter & B. Maro *Nature* 442, doi:10.1038/nature04907 (2006)

Hiiragi *et al.*[1] compare our model[2] of the developing mouse egg with theirs[3]. They seem to present patterning as equivalent to determination, but this is confusing as patterning does not have to mean determination. We have never stated that mouse embryo development is determined. Mouse development is regulative rather than determinative, and this can be explained in two ways: first, development could be entirely unbiased, generating identical cells; second, there could be some developmental bias or, in other words, pattern from the beginning that is not constraining. There is evidence for early bias from several laboratories[2,4–8] but this does not mean that cells have localized determinants fixing their fates. Regulative development does not exclude bias, which indicates inclination, not determination.

In a regulative system, where it is important to avoid invasive experimentation, technical differences between groups can be significant. We have tried but failed to confirm Hiiragi and Solter's rule that first cleavage always occurs perpendicular to the plane of pronuclei alignment[3] and now respond to the points raised by Hiiragi *et al.*[1].

Our time-lapse studies, which used multiple focal planes to monitor position carefully, indicated that cleavage tends to correlate with polar body and sperm entry sites before any flow of membrane to the cleavage furrow[2]. This would be difficult to find by studying just a single focal plane as others have done[3,9]. We note that their studies of cleavage did not use 40–80 sections, as the reference to a separate study of blastocysts[10] might imply, but just a single section. Also, our findings were similar for wild-type and H2B–EGFP strains and the conditions used for culture and microscopy allowed normal development[2]. In contrast to Hiiragi and Solter[3], we applied no selection or manipulation/alignment, but observed cleavage in all embryos in multiple planes[2].

We observed cleavage between pronuclei, as described by Hiiragi[3], only when we interfered with egg shape or disrupted the actin cytoskeleton[2]. Thus, it is possible that Hiiragi and co-workers inadvertently preselected embryos of a particular shape when aligning them because, in our experience, only some remain positioned on the culture dish with the polar body and the two pronuclei in the same plane. This could bias their data set as cell shape has an overriding effect[11]. To guard against such bias, we monitored all embryos and our findings[2] did not support their conclusions[3]. Furthermore, Hiiragi and Solter's model of cleavage states that female and male chromatin mix on the first metaphase plate[3], which, to the best of our knowledge, disagrees with the experimental observations of all other groups (see ref. 12, for example).

Our findings that progeny of early blastomeres make a biased contribution to different blastocyst parts are consistent with independent findings[4,5,7]. In assessing this, we found that it was critical to count the proportion of cells contributing to different blastocyst parts: to be significant, this had to be more than 70%[13], and not simply three or fewer cells[1].

Most important, if cleavage were random and blastomere fate could not be predicted, as Hiiragi *et al.* suggest, we would not have been able to isolate four-cell-stage blastomeres of like identity from different embryos and combine them to make chimaeras with different developmental properties[8]. Further weaknesses in the case against prepatterning are discussed elsewhere[14].

Hiiragi *et al.*[1] restate a model to account for how lineages of cells are established. Our findings that both two-cell blastomeres contribute to the inner cell mass and the trophectoderm[6,13], and that these lineage decisions are taken from the eight-cell stage onwards, support this long-standing model. Thus, we contend that lineage establishment is not the topic under discussion but that the issue is whether the cavitation site, and therefore the embryonic–abembryonic axis orientation, is defined stochastically or not. Our results indicate that early cleavage patterns can bias the site of cavity formation (Fig. 1).

The controversy seems in part to be semantic, resting on the interpretation of prepatterning and determination. By prepatterning, we mean some developmental bias. It does not connote determination: indeed, our own experiments to remove parts of the egg surgically indicate that determinants do not exist[15].

Berenika Plusa*, Anna-Katerina Hadjantonakis†, Dionne Gray*, Karolina Piotrowska-Nitsche*, Agnieszka Jedrusik*, Virginia E. Papaioannou‡, David M. Glover*, Magdalena Zernicka-Goetz*
*Gurdon Institute and Department of Genetics, University of Cambridge, Cambridge CB2 1QR, UK
e-mail: mzg@mole.bio.cam.ac.uk
†Developmental Biology Program, Sloan-Kettering Institute, New York, New York 10021, USA
‡Department of Genetics and Development, College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA

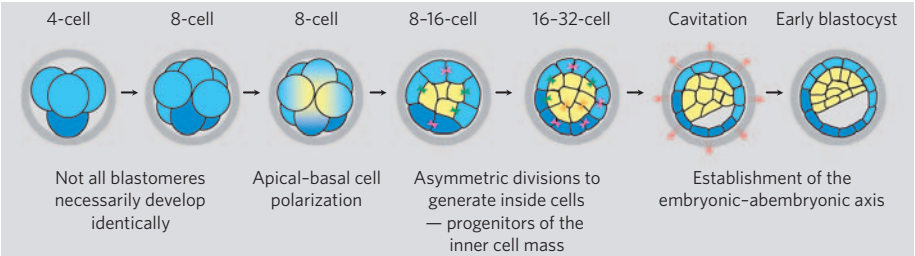


Figure 1 | Differences between blastomeres arise as a result of different cleavage order and orientation; differences accumulate and bias the positioning of cells and hence their fate. In this model, progeny of one cell (dark blue) tend to undertake more symmetric than asymmetric divisions from the eight-cell stage onwards. If the outside cells resulting from symmetric divisions are more weakly connected to inside cells, this would create a preferred site at which fluid could accumulate when the cavity forms. Inside cells, yellow; outside cells, blue; red or green links connect sister cells after symmetric or asymmetric division of outside cells, respectively; orange links connect sister inside cells; orange arrows indicate expansion of the blastocyst as the cavity expands. Not all sister-cell relationships are shown, as sisters can occupy different planes of the spherical embryo. Inside cells give rise to pluripotent inner cell mass (progenitors for future body and some extra-embryonic tissues); outside cells give rise to trophectoderm (progenitors for extra-embryonic tissues).

1. Hiiragi, T., Louvet-Vallée, S., Solter, D. & Maro, B. *Nature* **442,** doi:10.1038/nature04907 (2006).
2. Plusa, B. *et al. Nature* **434,** 391–395 (2005).
3. Hiiragi, T. & Solter, D. *Nature* **430,** 360–364 (2004).
4. Gardner, R. L. *Development* **124,** 289–301 (1997).
5. Gardner, R. L. *Development* **128,** 839–847 (2001).
6. Piotrowska, K., Wianny, F., Pedersen, R. A. & Zernicka-Goetz, M. *Development* **128,** 3739–3748 (2001).
7. Fujimori, T., Kurotaki, Y., Miyazaki, J. I. & Nabeshima ,Y. I. *Development* **21,** 5113–5122 (2003).
8. Piotrowska-Nitsche, K., Perea-Gomez, A., Haraguchi, S. & Zernicka-Goetz, M. *Development* **132,** 479–490 (2005).
9. Louvet-Vallée, S., Vinot, S. & Maro, B. *Curr. Biol.* **15,** 464–469 (2005).
10. Motosugi, N., Bauer, T., Polanski, P., Solter, D. & Hiiragi, T. *Genes Dev.* **19,** 1081–1092 (2005).
11. Gray, D. *et al. Curr. Biol.* **14,** 397–405 (2004).
12. Mayer, W., Smith, A., Fundele, R. & Haaf, T. *J. Cell Biol.* **148,** 629–634 (2000).
13. Piotrowska-Nitsche, K. & Zernicka-Goetz, M. *Mech. Dev.* **122,** 487–500 (2005).
14. Gardner, R. L. *Reprod. Biomed. Online* **12,** 144–149 (2006).
15. Zernicka-Goetz, M. *Development* **125,** 4803–4808 (1998).

doi:10.1038/nature04908

© 2006 Nature Publishing Group

REVIEWS

Clarifying the mechanics of DNA strand exchange in meiotic recombination

Matthew J. Neale[1] & Scott Keeney[1]

During meiosis, accurate separation of maternal and paternal chromosomes requires that they first be connected to one another through homologous recombination. Meiotic recombination has many intriguing but poorly understood features that distinguish it from recombination in mitotically dividing cells, and several of these features depend on the meiosis-specific DNA strand exchange protein Dmc1 (disrupted meiotic cDNA1). Many questions about this protein have arisen since its discovery more than a decade ago, but recent genetic and biochemical breakthroughs promise to shed light on the unique behaviours and functions of this central player in the remarkable chromosome dynamics of meiosis.

Sexual organisms must halve the chromosome number in gametes to maintain genome size with each generation. This goal is achieved through meiosis, in which two rounds of chromosome segregation follow a single round of DNA replication. The first meiotic division separates maternal and paternal copies of each chromosome, but in order for this segregation to occur properly, the chromosomes must first pair with their correct partner and then become physically connected so that they orient together on the meiotic spindle. Connection is established by the exchange of homologous chromosome arms (the point of crossover being called a chiasma) plus cohesion between sister chromatids[1,2] (Fig. 1). Exchange uses a specialized pathway of homologous recombination that repairs DNA double-strand breaks (DSBs)[3]—basically, the cell deliberately damages its own DNA and then uses the repair process to lock homologous chromosomes together. A central step in recombination involves proteins related to bacterial RecA that catalyse the pairing and exchange of DNA strands between single-stranded DNA (ssDNA) formed at the break and intact, homologous double-stranded DNA (dsDNA)[4]. Most eukaryotes have two RecA homologues, the ubiquitous Rad51 and its meiosis-specific counterpart Dmc1 (ref. 5).

Meiotic recombination is subject to many layers of control that dictate, for example, the choice of DNA substrate for DSB repair, the distribution and timing of recombination events, and the integration of recombination with higher order chromosome structures[2,6]. Executing these controls requires meiosis-specific factors (such as Dmc1) in addition to proteins (such as Rad51) that function during normal repair of DSBs in other cell types. The widely conserved Dmc1 has a critical role in meiosis in most sexual organisms, but what its molecular functions are and how and why it differs from its cousin Rad51 are not well understood[5]. Recent genetic studies have uncovered new information about the roles of Dmc1 and accessory proteins that modulate Dmc1 function *in vivo*. These advances, coupled with breakthroughs in biochemical analysis of Dmc1 activities, promise to answer long-standing questions about the mechanism of meiotic recombination.

The challenges of meiotic recombination

The molecular events of meiotic recombination have been most thoroughly described in the budding yeast *Saccharomyces cerevisiae* (Fig. 2). DSBs are formed by the meiosis-specific Spo11 protein, which cuts DNA via a topoisomerase-like reaction to generate covalent protein–DNA linkages to the 5′ DNA ends on either side of the break[3]. After Spo11 is removed from the DNA ends, one or more exonucleases process the DSB to generate 3′ ssDNA tails. DNA strand exchange proteins bind these tails, forming helical nucleoprotein filaments, which are the active structures that carry out the search for a homologous target and catalyse strand exchange. The detailed biochemical steps in this process are described below. Further processing of the strand exchange intermediates yields intact recombinant products that either have exchanged the flanking DNA arms (crossovers) or have not undergone exchange (non-crossovers) (Fig. 2).

The task of DNA strand exchange proteins in meiosis is remarkable. Each DSB generates about 1 kilobase of ssDNA, for which the homologous target duplex must be located and engaged. In yeast, this is the equivalent of searching through a 10-km-long string to find a particular 20-cm-long region. In human cells, that 20-cm stretch would need to be picked out of some 2,500 km, enough to reach from

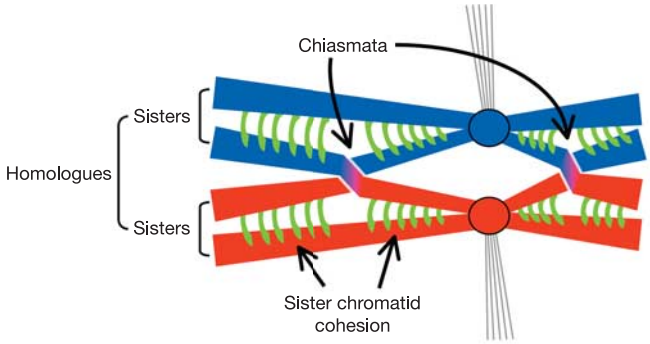


Figure 1 | Connections formed between homologous chromosomes during meiosis. Meiotic cells of most sexual organisms contain two copies of most chromosomes, one from each of the parents (red and blue). After DNA replication, each chromosome comprises a pair of sister chromatids held together by cohesion complexes (green). Sister centromeres (circles) attach as a single unit to microtubules (thin lines) from a spindle pole (not shown). Exchange of chromosome arms between non-sister chromatids yields a chiasma. Dissolution of sister chromatid cohesion along the arms allows the homologues to separate at the first meiotic division (not shown).

[1]Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA.

© 2006 Nature Publishing Group

London to Moscow. The problem is more complex, however, because within that 2,500-km length would be three exact matches to the target (one on the sister chromatid, and one on each of the two chromatids of the homologous chromosome), but only the two targets on the homologue can be used. This strong preference for interhomologue recombination in meiosis is important because intersister recombination would not join homologues. This bias contrasts with mitotic cells, in which the sister chromatid is greatly preferred as the target for Rad51-mediated recombination[7]. The problem gets even worse because every meiotic cell makes many DSBs (hundreds per cell in yeast and mouse)[8,9]. Thus, repair has to occur simultaneously at numerous places in each meiosis. Finally, the search for homology and the catalysis of strand exchange must be spatially and temporally coordinated with the development of higher order chromosome structures (for example, synaptonemal complex, chromosome axes and chromatin loops) that are required for the recombination event to successfully connect the homologues[2,6]. All of this is accomplished in just a few hours in yeast and a few days in mammals. With all these issues to consider, it is clear why meiotic recombination is so tightly controlled, and why this process has fascinated chromosome biologists for decades.

The bacterial DNA strand exchange protein RecA had already been extensively characterized[4] when the first eukaryotic homologues were identified in *S. cerevisiae*[10–12]. The widely conserved Rad51 protein gets its name from the radiation-sensitive phenotype caused by mutations in the gene. Not surprisingly then, Rad51 is required for most homologous recombination reactions in both mitotic and meiotic cells[13]. In contrast, Dmc1 is meiosis-specific and, in many organisms, *dmc1* mutants display reduced or absent meiotic recombination[5]. It was appreciated early on that Dmc1 activities must be central to the unique aspects of meiotic (as opposed to mitotic) recombination[10,14]. As such, several fundamental questions have stood since its discovery: which aspects of meiotic recombination reflect specialized roles for Dmc1? Is it a DNA strand exchange protein, and what are the intrinsic biochemical properties that differentiate it from Rad51? What are the extrinsic factors that control and direct Dmc1, and to what extent do these dictate the unique properties of Dmc1?

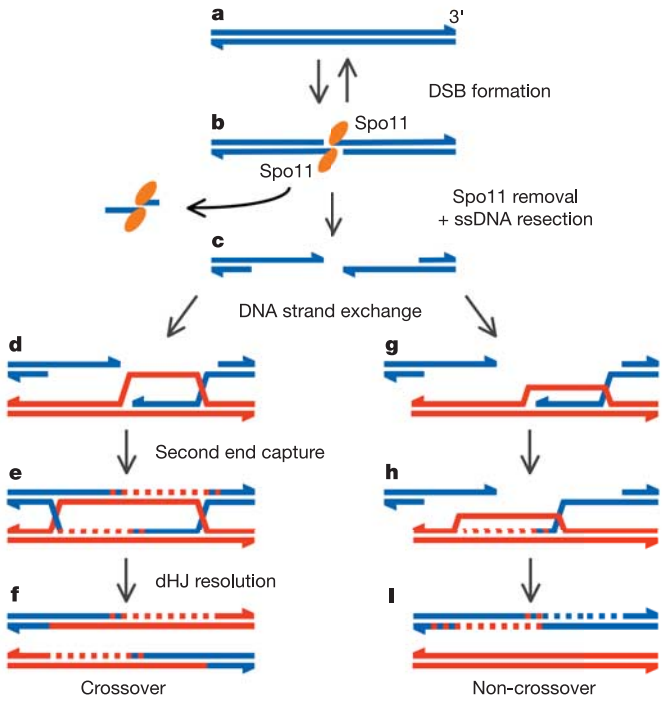


Figure 2 | DNA events in meiotic recombination. a–c, Presynapsis. Spo11 (elipses) cleaves dsDNA, yielding a covalent Spo11–DNA complex. Endonuclease releases Spo11 bound to a short oligonucleotide, and 5′ DNA strands are degraded to yield 3′ ssDNA tails, which are bound by Rad51 and Dmc1 (not shown). **d–f**, Crossover formation. **d**, Invasion of ssDNA from one end of the break forms an asymmetric strand exchange intermediate. **e**, DNA synthesis (dashed line) is primed from the invading 3′ end; the second DSB end is captured and primes DNA synthesis. Ligation yields a pair of Holliday junctions (dHJ). **f**, Resolution yields a mature product with exchanged flanking DNA. **g–i**, A non-crossover pathway. Strand invasion (**g**) and DNA synthesis (**h**) are inferred but have not been directly detected. A transient strand invasion complex may be dissociated, perhaps by DNA helicases, allowing newly synthesized DNA to anneal to complementary ssDNA on the other side of the break (**i**). Further DNA synthesis and ligation yield a mature non-crossover product.

Genetics and cytology of Dmc1

The physiological function and specialized roles of Dmc1 have been best characterized in *S. cerevisiae*[5]. In one commonly used laboratory strain (the SK1 background), *dmc1* mutants show a near-complete block to recombination[10]. When the DNA intermediates of recombination were assayed, no stable strand invasion or double-Holliday-junction formation was detected[14,15] (Fig. 2). Based in part on these observations, it was suggested that Dmc1 promotes an interhomologue-only recombination pathway that is unique to meiosis[14]; this role was also suggested by the observation that *dmc1* mutants have increased recombination between ectopic sequence repeats (that is, homologous sequences located at positions other than the equivalent position on the homologous chromosome)[16]. *rad51* mutants are similarly profoundly defective for meiotic recombination, revealing that both DNA strand exchange proteins contribute critical, non-overlapping functions[5]. In every organism examined, Rad51 and Dmc1 are components of multiprotein complexes, or nodules, that form on meiotic chromosomes at the sites where DSBs are being repaired[8,9].

A striking hallmark of meiotic recombination in most organisms is that crossovers are not randomly distributed along chromosomes. Instead, the presence of a crossover makes it less likely that another will form nearby[17]. This phenomenon, known as crossover interference, is poorly understood, but Dmc1 may be involved because interference is lost or diminished when Dmc1 is absent[18,19].

Interestingly, the 5′ strands of DSBs are more extensively processed by exonuclease(s) when Dmc1 and/or Rad51 is missing[10,11,14]. This finding indicates that the nuclease(s) and DNA strand exchange proteins are coordinated in some fashion, for example through competition for the same substrate or by direct protein–protein interaction. This coordination probably serves to strike a balance between too little processing (and thus, not enough ssDNA to carry out the homology search and strand exchange) and too much processing (which may be deleterious for integrity of the genome). The idea of a hand over from the DSB processing activity to the strand exchange proteins is attractive because functional coupling of sequential steps in the pathway might help meiotic DSB repair to operate so efficiently that it can handle larger numbers of DSBs than would be tolerated by a mitotic cell.

Dmc1 is important for meiotic recombination in many organisms; for example, mice with targeted mutation of the *Dmc1* gene are sterile and show hallmarks of poorly repaired DSBs[20]. However, Dmc1 is not absolutely essential in all circumstances. Some organisms, such as *Drosophila melanogaster*, *Caenorhabditis elegans* and *Neurospora crassa*, have Rad51 but lack a Dmc1 orthologue[17]. In another frequently used *S. cerevisiae* laboratory strain (the BR background), and in *Schizosaccharomyces pombe*, *dmc1* mutants still exhibit significant levels of meiotic recombination[21,22]. Moreover, the recombination defect in *S. cerevisiae dmc1* mutants can be largely suppressed by overexpression of Rad51 or Rad54 (a protein that stimulates Rad51; see below)[19,23]. These findings may indicate that Rad51 and Dmc1 operate in multiple meiotic recombination pathways that only partially overlap in budding yeast[19]. However, it is also possible that under normal circumstances the two proteins function only together, and that only aberrant recombination occurs when one is missing.

© 2006 Nature Publishing Group

Dmc1-mediated DNA strand exchange

Whereas *in vivo* studies identify aspects of meiotic recombination that are influenced by Dmc1, they do not reveal the biochemical mechanism of Dmc1 action. Given its homology to RecA, it was long suspected that Dmc1 would have DNA strand exchange activity[9,24], but only relatively recently has this been demonstrated. Strand exchange by RecA-related proteins can be divided into three phases[4] (Fig. 3a). During the presynaptic phase, the strand exchange protein assembles on ssDNA, creating a helical nucleoprotein filament in which the DNA is stretched and underwound relative to B-form DNA. Formation of a ternary complex of the ssDNA–protein filament with dsDNA initiates the synaptic phase. Juxtaposition of three DNA strands within the synaptic filament permits rapid homology sampling through transient Watson–Crick base pairing between the ssDNA and the complementary strand of the duplex partner. Sequential cycles of binding, sampling and release of dsDNA are performed until homology is detected. A stable DNA joint is then formed by intertwining of the ssDNA with its complement from the homologous target. During the postsynaptic phase, the exchange of DNA strands is completed and the filament is dismantled. The generation of mature recombinant products (that is, intact DNA duplexes with or without reciprocal exchange of flanking sequences) requires many other postsynaptic processes such as more extensive strand exchange via migration of branched DNA structures (D loops or Holliday junctions); resolution of branched DNA intermediates by helicases, nucleases and/or topoisomerases; DNA synthesis; and resealing of DNA strand interruptions with DNA ligase[13,25].

Two types of *in vitro* DNA strand exchange assay are often used. The first detects transfer of one strand of a linear dsDNA onto a circular single-stranded nucleoprotein complex, creating a nicked duplex circle and a linear ssDNA (Fig. 3b). (Note that this situation contrasts with recombination *in vivo*, where a 3′ DNA end is provided by the initiating ssDNA.) The second assay detects invasion of a short ssDNA nucleoprotein complex into a negatively supercoiled duplex circle, displacing the non-complementary strand into a D-loop structure (Fig. 3c). Because RecA has high specificity for binding ssDNA over dsDNA, strand exchange is efficient even when all reaction components are mixed simultaneously[4]. In contrast, Dmc1 is more sensitive to the order of addition because its ability to bind dsDNA can lead to non-productive complexes being formed[26].

In early studies, Dmc1 displayed poor strand exchange activity relative to RecA or Rad51 (refs 27–30). Although RecA and its other homologues (eukaryotic Rad51, bacteriophage UvsX and archaeal RadA) form hexameric to octameric rings in solution, the active oligomeric form for strand exchange is a helical nucleoprotein filament[4]. Electron microscopy of Dmc1, however, revealed nucleoprotein complexes that were exclusively rings, never helices[29,31] (Fig. 4a). The lack of helical filaments was surprising because Dmc1 has high sequence conservation with RecA-related proteins and because biophysical studies revealed that Dmc1 and Rad51 recognize homology in the same way[32]. It was suggested that rings were an artefact caused by lack of essential cofactors[32], but models for ring-based DNA strand exchange were also proposed[33].

More recently, a robust Dmc1-mediated strand exchange reaction has been described, which differs from previous studies in using physiological pH and higher salt concentrations[26,34–36]. Under these conditions, helical Dmc1–ssDNA filaments were formed[26,35,37] (Fig. 4b), strongly indicating that the filament is the active oligomeric form for DNA strand exchange. Inefficient filament formation probably accounts for the weak strand exchange activity of earlier studies.

Yet, is there a function for the Dmc1 ring? It has been proposed that stacks of Dmc1 rings on DNA might be converted into an active helix, perhaps upon binding ATP[38]. As Dmc1 rings are oriented with alternating head-to-tail polarity[31,33], however, it is impossible to convert a stack of rings into a continuous helical filament without dismantling and reorienting at least half of the DNA-bound protein. A more likely possibility is that free, but not DNA-bound, rings are the form of Dmc1 (and of other RecA relatives) that assembles on DNA (Fig. 4c).

Ca^{2+} stimulates both filament formation on ssDNA and DNA strand exchange by bacterially expressed human and yeast Dmc1 proteins[35,36]. Ca^{2+} also stimulates Rad51 (ref. 39), so it has been suggested that Dmc1 and Rad51 activity might be modulated *in vivo* by variation in intracellular calcium levels[35,36]. However, the free Ca^{2+} concentration required for optimal Dmc1 stimulation (~100 μM) is much higher than the free concentration in the cytosol (0.1–1 μM in yeast), and Ca^{2+} was not necessary for efficient DNA strand exchange by human DMC1 expressed in insect cells and assayed in higher salt concentrations[26]. Therefore it is possible that Ca^{2+} acts as a non-physiological stimulant, perhaps substituting for more

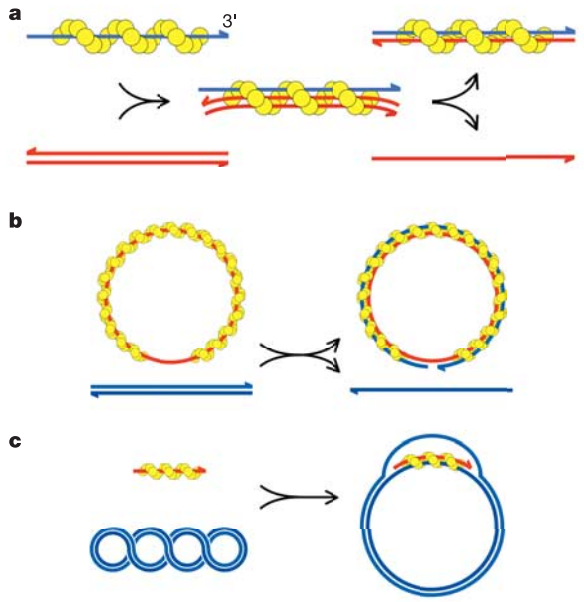


Figure 3 | The DNA strand exchange reaction. **a**, Three stages of strand exchange: presynaptic helical filament on ssDNA (left); synaptic complex with homologous dsDNA (middle); postsynaptic stage with the original ssDNA fully base-paired with the complementary strand from the dsDNA donor (right). **b**, **c**, Typical *in vitro* DNA strand exchange assay systems. **b**, Strand transfer from a linear duplex to a circular nucleoprotein filament. **c**, D-loop assay. Invasion of a short nucleoprotein filament into a negatively supercoiled circular duplex, creating a bubble, or D-loop.

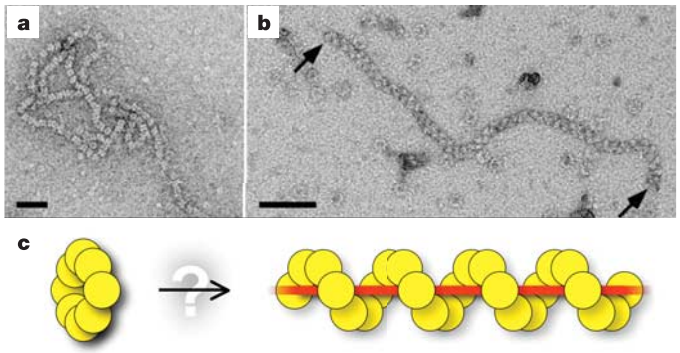


Figure 4 | Oligomeric structure of Dmc1 nucleoprotein complexes. **a**, **b**, Electron micrographs of Dmc1–DNA complexes (from refs 26, 31, with permission). Scale bars, 50 nm. **a**, Stacks of octameric Dmc1 rings on DNA. **b**, A helical Dmc1–ssDNA filament, similar to structures created by other RecA family members. **c**, A Dmc1 ring and a right-handed helical nucleoprotein filament. The dynamic relationship between rings and helical filaments is not known, but it is possible that filaments are formed by the opening and deposition of Dmc1 rings.

© 2006 Nature Publishing Group

physiological monovalent cations and/or mimicking the effects of accessory proteins.

DNA strand exchange proteins RecA, Rad51 and Dmc1 require nucleotide cofactor, but whether Dmc1 requires ATP hydrolysis remains controversial. Some studies indicate a strong requirement for ATP hydrolysis[26,28,34], whereas others show strand exchange even when hydrolysis is hindered[27,35,36,40–42]. Notably, neither RecA nor Rad51 requires ATP hydrolysis for strand invasion, but both require energy to turn over and release product DNA[4].

The flurry of recent papers describing robust Dmc1-catalysed DNA strand exchange represent an important landmark because they open the door to a more detailed mechanistic understanding of Dmc1. The key question is, what are the intrinsic biochemical differences that functionally distinguish Dmc1 from Rad51? One apparent difference is the greater sensitivity of Dmc1 to reaction conditions *in vitro*, which seems to be related, at least in part, to the dynamics of ring versus filament formation. At the moment, it is not obvious if or how these properties relate to the unique requirements of meiotic recombination. The challenge will be to sort out whether these are meaningful differences (as opposed to by-products of suboptimal *in vitro* conditions), and if so, to translate the differences into functional consequences *in vivo*.

Modulation of Dmc1 activity by extrinsic factors

RecA family members are remarkable enzymes, but do not function alone—RecA and Rad51 are profoundly affected by accessory proteins *in vivo* and *in vitro*[13,43]. These accessory factors can function at one or more steps (presynaptic, synaptic, postsynaptic), and they serve to enhance (or sometimes restrain) catalytic efficiency and/or appropriately target strand exchange activities. Similarly, Dmc1 is also highly modulated[5]. Understanding how these factors affect Dmc1 is critical to understanding the mechanisms of the extremely regulated pathways of meiotic recombination.

RPA and recombination mediators. The ssDNA-binding protein RPA is required for efficient strand exchange by Rad51 but its effect is complex[43]. RPA minimizes secondary structure in ssDNA, promoting Rad51 filament formation; it also sequesters both free ssDNA substrate and the displaced single strand after strand exchange, preventing either from competing with dsDNA. However, RPA can also directly compete with Rad51 for binding to ssDNA, which is inhibitory. RPA is required for extensive Dmc1-catalysed strand exchange, presumably for the same reasons as for Rad51 (ref. 26).

In vivo, RPA probably competes with Dmc1 for binding to ssDNA at meiotic DSBs. For Rad51-dependent reactions *in vitro*, inclusion of Rad52 or a complex of Rad55 and Rad57 overcomes this inhibitory effect and promotes assembly of Rad51 filaments on RPA-coated ssDNA[13,43]. Proteins such as these are generally referred to as recombination 'mediators'[43]. It is not yet known if they directly stimulate Dmc1 as well, but all are involved in meiotic recombination (at least in some organisms)[13]. The tumour suppressor protein Brca2 has also been implicated as a Rad51 mediator[44]. In many organisms, Brca2-deficient mutants have meiotic recombination defects[45–48] and, in plants at least, Brca2 and Dmc1 interact physically[46,49].

Rdh54. Genetic studies in yeast indicate an important role in meiotic recombination for Rdh54 (also known as Tid1), a homologue of the DNA repair protein Rad54. Rdh54 is required for the timely conversion of DSBs into mature recombinant products[50]. In its absence, there is decreased co-localization of Dmc1 and Rad51 in cytological nodules on chromosomes[51], presumably reflecting a role for Rdh54 in dissociating Dmc1 from duplex DNA not associated with DSBs (D. K. Bishop, personal communication). Rad54 is less important in the meiotic context in yeast, so the role of Rdh54 is distinct[50,51]. Rdh54 promotes D-loop formation by Rad51 *in vitro*[52,53], so it may function similarly with Dmc1. Fission and budding yeast Rdh54 and Dmc1 interact[54,55]. Mammals have two known Rad54 family members, Rad54 and Rad54B. Human RAD54B interacts with and stimulates DMC1 *in vitro*[26], but the significance of this interaction is not yet clear because there is little or no meiotic recombination defect in mice lacking Rad54B, Rad54, or both[56].

Insight into possible biochemical functions of Rdh54 with Dmc1 comes from consideration of its relative, Rad54, a member of the SWI2/SNF2 family of ATPases, which appears to stimulate Rad51 activity through multiple mechanisms[13,43]. Rad54 couples ATP hydrolysis with translocation along the DNA, thereby generating superhelical torsion in the DNA[57,58]. Rad51 stimulates this activity, and the topological changes in the DNA in turn promote the ability of Rad51 to catalyse strand invasion and D-loop formation[57,58]. Rad54 also possesses chromatin remodelling activity, stimulated by Rad51, indicating that another Rad54 function is to enhance the ability of Rad51 to target chromatinized dsDNA for strand exchange[59,60]. Finally, Rad54 dismantles Rad51 complexes bound to dsDNA[61]; this activity may be useful postsynaptically to recycle limiting amounts of Rad51.

Hop2 and Mnd1. Genetic studies in yeast and mouse have identified a chromosome-associated, heterodimeric complex of Hop2 and Mnd1 proteins. Loss of these proteins causes chromosomes to synapse non-homologously and meiotic DSBs to persist[40,62–67]. Purified human, mouse and budding yeast Hop2–Mnd1 complexes all stimulate D-loop formation and, where tested, DNA strand exchange by the Dmc1 proteins from the same organism[40,42,68]. The mammalian complexes also stimulate Rad51 (refs 42, 68). Whether yeast Hop2–Mnd1 affects Rad51 activities has not been reported, but *hop2* and *mnd1* mutant phenotypes are much closer to those of *dmc1* mutants than *rad51*, leading to the hypothesis that Hop2–Mnd1 functions primarily with Dmc1 (refs 40, 65, 66). Indeed, organisms that do not contain a Dmc1 orthologue also lack Hop2 and Mnd1 (ref. 17). However, because Rad51 and Dmc1 are so closely tied together (see below), it is plausible that Hop2–Mnd1 could be specific for a Dmc1-dependent pathway and yet function biochemically to stimulate both Rad51 and Dmc1.

The molecular function of Hop2–Mnd1 remains unclear. The biochemical studies are consistent with direct effects on the catalytic activities of Dmc1 and/or Rad51. A presynaptic role in aiding the binding of Dmc1 to ssDNA has been suggested[42], but the genetics and biochemistry are currently most consistent with a synaptic and/or postsynaptic role[40,42,64,66,67]. Klein and colleagues have suggested a different model, in which Hop2–Mnd1 promotes Dmc1 activity indirectly by acting on the chromatin and/or higher order chromosome structures of the homologous target[66]. This model is motivated in part by the observation that Hop2–Mnd1 associates with chromatin independent of DSB formation, and does not co-localize with DSB sites or with Rad51 and Dmc1 complexes[66].

Mei5 and Sae3. The evolutionarily conserved Mei5 and Sae3 proteins of budding yeast form DSB-dependent chromosomal complexes that co-localize with each other and with Dmc1 in a mutually dependent manner[69,70]. Dmc1, Mei5 and Sae3 physically interact, and Mei5 and Sae3 co-purify as a complex[69]. Unrepaired meiotic DSBs accumulate in *mei5* and *sae3* mutants, indicating a defect in strand exchange[69,70]. Rad51 foci form in the absence of Mei5, Sae3 or Dmc1, suggesting that the role of Mei5–Sae3 is Dmc1-specific[69,70]. *S. pombe* contains two Mei5 homologues, Swi2 and Sfr1, and one Sae3 homologue, Swi5. Unlike in budding yeast, these proteins interact with Rhp51 (the *S. pombe* Rad51 orthologue) and function in recombination in both vegetative and meiotic cells[71,72].

The biochemical function(s) of Mei5–Sae3 is not yet known. The fact that Dmc1 is not loaded onto chromosomes in *mei5* and *sae3* mutants (unlike in *hop2* and *mnd1* mutants) suggests a direct presynaptic role in promoting Dmc1 filament formation[69,70]. Perhaps Mei5–Sae3 converts inactive Dmc1 rings into active helical filaments, or helps to load Dmc1 onto RPA-coated ssDNA[69] (as Rad52, Rad55 and Rad57 perform for Rad51). Indeed, genetic experiments in *S. pombe* support a mediator role for Swi5–Sfr1 in Rhp51-mediated recombination[71]. Now that Mei5 and Sae3 can be purified[69], these ideas can be tested.

Rad51. The function of Dmc1 in meiotic recombination is closely

© 2006 Nature Publishing Group

coordinated with Rad51 (ref. 5). As noted above, both proteins localize together to the sites where DSBs are being repaired[8,9]. In the absence of Rad51, localization of yeast Dmc1 is impaired, although Rad51 localization seems normal in the absence of Dmc1 (refs 9, 24). Precisely how the proteins are coordinated is not yet clear. Based in part on co-localization experiments, it has been proposed that they work together by forming mixed nucleoprotein filaments[8]. Homopolymeric Rad51 and Dmc1 filaments arranged consecutively on the same ssDNA is another possibility. On the other hand, side-by-side immunostaining patterns have been observed for the two proteins under at least some circumstances on spread meiotic chromosomes[8,51], leading to the hypothesis that Rad51 and Dmc1 bind to opposite ends of each DSB[15,51]. The idea of asymmetrical protein binding is attractive because the two ends of a DSB themselves behave differently—one end carries out the initial strand invasion, while the other end is captured in a separate reaction[15] (see Fig. 2). If this idea is correct, it raises a couple of interesting questions: which protein is responsible for initial strand invasion, and how is the asymmetric distribution of the proteins accomplished? It has been proposed that the Dmc1 nucleoprotein filament is the end that invades first[15], although this remains to be verified. Recent studies of the early steps in DSB processing, when Spo11 is removed from its covalent attachment to DNA ends, have led to the hypothesis that asymmetry is set up very early, at or before DSB formation[73]. This also awaits confirmation.

Structural components of meiotic chromosomes. Meiotic recombination and large-scale chromosome structures are highly integrated, in part because a functional chiasma itself is a multicomponent structure comprising local exchange of homologous DNA duplexes (that is, recombination) plus exchange of the proteinaceous axes of the homologues and local separation of sister chromatids[2,6] (Fig. 1). Achieving this integration requires cross-talk spanning size scales that differ by orders of magnitude; that is, between the enzymes executing the DNA events of recombination and the structural proteins that make up micrometre-scale chromosome structures. The detailed molecular mechanisms involved are poorly understood. An extensive discussion is beyond the scope of this review (for further information see refs 1, 2, 6, 74), but a few points pertaining specifically to Dmc1 will be mentioned here.

The Red1 protein in *S. cerevisiae* provides one type of functional connection between Dmc1 and chromosome organization. Red1 is a major structural component of meiotic chromosomes[75]. *In vivo*, one function of Red1 is to promote, directly or indirectly, the loading of Dmc1 onto ssDNA at meiotic DSBs[74]. Another role of Red1 is to prevent stable DNA strand exchange between homologues when Dmc1 is absent[14,15]. Whereas DSBs persist in *dmc1* mutant cells, cells lacking both Dmc1 and Red1 are able to repair meiotic DSBs by Rad51-mediated recombination between sister chromatids instead of between homologues. Thus, it seems that interplay between Dmc1 and Red1 is integral to the interhomologue bias of meiotic recombination.

Another connection between Dmc1 and chromosome structure is through Rec8, an evolutionarily conserved, meiosis-specific subunit of cohesin, the multiprotein complex that holds sister chromatids together in both mitosis and meiosis[1,76]. In addition to defects in sister chromatid cohesion, *rec8* mutants in *S. cerevisiae* have recombination phenotypes similar in some respects to *dmc1* mutants, including inefficient repair and over digestion by exonucleases[76]. Recombination defects are also apparent in a *Rec8*-defective mouse, suggesting that at least some of the roles of this protein in recombination are conserved[77].

Outlook

This is an exciting time for students of meiotic chromosome dynamics in general, and of meiotic recombination in particular. The field faces several challenges. On the biological front, there is a need for more information about the specific processes influenced by Dmc1 (for example, interhomologue bias, crossover interference), particularly in organisms other than budding yeast. Which aspects of Dmc1 function are conserved, and which are not? How does recombination differ in organisms that have Dmc1 versus those that do not? On the biochemical front, there is a need for deeper mechanistic understanding of the mediators and other accessory proteins that influence Dmc1 activities. How do they modify Dmc1 activity? What is the interplay between Rad51 and Dmc1, and how do other factors impinge?

The biggest challenge, however, will be to connect the dots between the biochemical properties of Dmc1 and the *in vivo* functions revealed by genetics. One route to this goal will be to develop *in vitro* reactions that recapitulate increasingly complex aspects of chromosome behaviour, such as the connection between Dmc1 and chromosome structure proteins. Another route will be to apply the powerful genetic and cytological methods available in many organisms to test specific predictions from biochemical experiments, for example by examining phenotypes of mutants with specific biochemical defects in Dmc1 or other proteins. At such a point, we will be able to say that these too often disparate fields—biochemistry and genetics—have finally (re)combined.

1. Petronczki, M., Siomos, M. F. & Nasmyth, K. Un menage a quatre: the molecular biology of chromosome segregation in meiosis. *Cell* **112**, 423–440 (2003).
2. Kleckner, N. Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex. *Chromosoma* **115**, 175–194 (2006).
3. Keeney, S. Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol.* **52**, 1–53 (2001).
4. Bianco, P. R., Tracy, R. B. & Kowalczykowski, S. C. DNA strand exchange proteins: a biochemical and physical comparison. *Front. Biosci.* **3**, D570–D603 (1998).
5. Shinohara, A. & Shinohara, M. Roles of RecA homologues Rad51 and Dmc1 during meiotic recombination. *Cytogenet. Genome Res.* **107**, 201–207 (2004).
6. Zickler, D. & Kleckner, N. Meiotic chromosomes: integrating structure and function. *Annu. Rev. Genet.* **33**, 603–754 (1999).
7. Johnson, R. D. & Jasin, M. Double-strand-break-induced homologous recombination in mammalian cells. *Biochem. Soc. Trans.* **29**, 196–201 (2001).
8. Tarsounas, M., Morita, T., Pearlman, R. E. & Moens, P. B. RAD51 and DMC1 form mixed complexes associated with mouse meiotic chromosome cores and synaptonemal complexes. *J. Cell Biol.* **147**, 207–220 (1999).
9. Bishop, D. K. RecA homologs Dmc1 and Rad51 interact to form multiple nuclear complexes prior to meiotic chromosome synapsis. *Cell* **79**, 1081–1092 (1994).
10. Bishop, D. K., Park, D., Xu, L. & Kleckner, N. *DMC1*: a meiosis-specific yeast homolog of *E. coli recA* required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* **69**, 439–456 (1992).
11. Shinohara, A., Ogawa, H. & Ogawa, T. Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **69**, 457–470 (1992).
12. Aboussekhra, A., Chanet, R., Adjiri, A. & Fabre, F. Semidominant suppressors of Srs2 helicase mutations of *Saccharomyces cerevisiae* map in the *RAD51* gene, whose sequence predicts a protein with similarities to procaryotic RecA proteins. *Mol. Cell. Biol.* **12**, 3224–3234 (1992).
13. Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38**, 233–271 (2004).
14. Schwacha, A. & Kleckner, N. Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. *Cell* **90**, 1123–1135 (1997).
15. Hunter, N. & Kleckner, N. The single-end invasion: an asymmetric intermediate at the double-strand break to double-Holliday junction transition of meiotic recombination. *Cell* **106**, 59–70 (2001).
16. Grushcow, J. M. *et al.* *Saccharomyces cerevisiae* checkpoint genes *MEC1, RAD17* and *RAD24* are required for normal meiotic recombination partner choice. *Genetics* **153**, 607–620 (1999).
17. Stahl, F. W. *et al.* Does crossover interference count in *Saccharomyces cerevisiae? Genetics* **168**, 35–48 (2004).
18. Shinohara, M., Sakai, K., Shinohara, A. & Bishop, D. K. Crossover interference in *Saccharomyces cerevisiae* requires a *TID1/RDH54*- and *DMC1*-dependent pathway. *Genetics* **163**, 1273–1286 (2003).
19. Tsubouchi, H. & Roeder, G. S. The importance of genetic recombination for fidelity of chromosome pairing in meiosis. *Dev. Cell* **5**, 915–925 (2003).
20. Di Giacomo, M. *et al.* Distinct DNA damage-dependent and independent responses drive the loss of oocytes in recombination-defective mouse mutants. *Proc. Natl Acad. Sci. USA* **102**, 737–742 (2005).
21. Rockmill, B., Sym, M., Scherthan, H. & Roeder, G. S. Roles for two RecA homologs in promoting meiotic chromosome synapsis. *Genes Dev.* **9**, 2684–2695 (1995).
22. Young, J. A., Hyppa, R. W. & Smith, G. R. Conserved and nonconserved proteins for meiotic DNA breakage and repair in yeasts. *Genetics* **167**, 593–605 (2004).
23. Bishop, D. K. *et al.* High copy number suppression of the meiotic arrest caused by a *dmc1* mutation: *REC114* imposes an early recombination block and *RAD54* promotes a *DMC1*-independent DSB repair pathway. *Genes Cells* **4**, 425–444 (1999).

© 2006 Nature Publishing Group

24. Shinohara, A., Gasior, S., Ogawa, T., Kleckner, N. & Bishop, D. K. *Saccharomyces cerevisiae recA* homologues *RAD51* and *DMC1* have both distinct and overlapping roles in meiotic recombination. *Genes Cells* **2,** 615–629 (1997).
25. Paques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63,** 349–404 (1999).
26. Sehorn, M. G., Sigurdsson, S., Bussen, W., Unger, V. M. & Sung, P. Human meiotic recombinase Dmc1 promotes ATP-dependent homologous DNA strand exchange. *Nature* **429,** 433–437 (2004).
27. Hong, E. L., Shinohara, A. & Bishop, D. K. *Saccharomyces cerevisiae* Dmc1 protein promotes renaturation of single-strand DNA (ssDNA) and assimilation of ssDNA into homologous super-coiled duplex DNA. *J. Biol. Chem.* **276,** 41906–41912 (2001).
28. Li, Z., Golub, E. I., Gupta, R. & Radding, C. M. Recombination activities of HsDmc1 protein, the meiotic human homolog of RecA protein. *Proc. Natl Acad. Sci. USA* **94,** 11221–11226 (1997).
29. Masson, J. Y. *et al.* The meiosis-specific recombinase hDmc1 forms ring structures and interacts with hRad51. *EMBO J.* **18,** 6552–6560 (1999).
30. Nara, T., Hamada, F., Namekawa, S. & Sakaguchi, K. Strand exchange reaction *in vitro* and DNA-dependent ATPase activity of recombinant LIM15/DMC1 and RAD51 proteins from *Coprinus cinereus*. *Biochem. Biophys. Res. Commun.* **285,** 92–97 (2001).
31. Passy, S. I. *et al.* Human Dmc1 protein binds DNA as an octameric ring. *Proc. Natl Acad. Sci. USA* **96,** 10684–10688 (1999).
32. Gupta, R. C., Golub, E., Bi, B. & Radding, C. M. The synaptic activity of HsDmc1, a human recombination protein specific to meiosis. *Proc. Natl Acad. Sci. USA* **98,** 8433–8439 (2001).
33. Kinebuchi, T. *et al.* Structural basis for octameric ring formation and DNA interaction of the human homologous-pairing protein Dmc1. *Mol. Cell* **14,** 363–374 (2004).
34. Sauvageau, S. *et al.* Fission yeast Rad51 and Dmc1, two efficient DNA recombinases forming helical nucleoprotein filaments. *Mol. Cell. Biol.* **25,** 4377–4387 (2005).
35. Lee, M. H. *et al.* Calcium ion promotes yeast Dmc1 activity via formation of long and fine helical filaments with single-stranded DNA. *J. Biol. Chem.* **280,** 40980–40984 (2005).
36. Bugreev, D. V., Golub, E. I., Stasiak, A. Z., Stasiak, A. & Mazin, A. V. Activation of human meiosis-specific recombinase Dmc1 by Ca^{2+}. *J. Biol. Chem.* **280,** 26886–26895 (2005).
37. Chang, Y. C. *et al.* Molecular visualization of the yeast Dmc1 protein ring and Dmc1–ssDNA nucleoprotein complex. *Biochemistry* **44,** 6052–6058 (2005).
38. Kinebuchi, T., Kagawa, W., Kurumizaka, H. & Yokoyama, S. Role of the N-terminal domain of the human DMC1 protein in octamer formation and DNA binding. *J. Biol. Chem.* **280,** 28382–28387 (2005).
39. Bugreev, D. V. & Mazin, A. V. Ca^{2+} activates human homologous recombination protein Rad51 by modulating its ATPase activity. *Proc. Natl Acad. Sci. USA* **101,** 9988–9993 (2004).
40. Chen, Y. K. *et al.* Heterodimeric complexes of Hop2 and Mnd1 function with Dmc1 to promote meiotic homolog juxtaposition and strand assimilation. *Proc. Natl Acad. Sci. USA* **101,** 10572–10577 (2004).
41. Enomoto, R. *et al.* Positive role of the mammalian TBPIP/HOP2 protein in DMC1-mediated homologous pairing. *J. Biol. Chem.* **279,** 35263–35272 (2004).
42. Petukhova, G. V. *et al.* The Hop2 and Mnd1 proteins act in concert with Rad51 and Dmc1 in meiotic recombination. *Nature Struct. Mol. Biol.* **12,** 449–453 (2005).
43. Sung, P., Krejci, L., Van Komen, S. & Sehorn, M. G. Rad51 recombinase and recombination mediators. *J. Biol. Chem.* **278,** 42729–42732 (2003).
44. Yang, H., Li, Q., Fan, J., Holloman, W. K. & Pavletich, N. P. The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA–ssDNA junction. *Nature* **433,** 653–657 (2005).
45. Sharan, S. K. *et al.* BRCA2 deficiency in mice leads to meiotic impairment and infertility. *Development* **131,** 131–142 (2004).
46. Siaud, N. *et al.* Brca2 is involved in meiosis in *Arabidopsis thaliana* as suggested by its interaction with Dmc1. *EMBO J.* **23,** 1392–1401 (2004).
47. Kojic, M., Kostrub, C. F., Buchman, A. R. & Holloman, W. K. BRCA2 homolog required for proficiency in DNA repair, recombination, and genome stability in *Ustilago maydis*. *Mol. Cell* **10,** 683–691 (2002).
48. Martin, J. S., Winkelmann, N., Petalcorin, M. I., McIlwraith, M. J. & Boulton, S. J. RAD-51-dependent and -independent roles of a *Caenorhabditis elegans* BRCA2-related protein during DNA double-strand break repair. *Mol. Cell. Biol.* **25,** 3127–3139 (2005).
49. Dray, E., Siaud, N., Dubois, E. & Doutriaux, M. P. Interaction between *Arabidopsis* Brca2 and its partners Rad51, Dmc1, and Dss1. *Plant Physiol.* **140,** 1059–1069 (2006).
50. Shinohara, M. *et al.* Characterization of the roles of the *Saccharomyces cerevisiae RAD54* gene and a homologue of *RAD54, RDH54/TID1*, in mitosis and meiosis. *Genetics* **147,** 1545–1556 (1997).
51. Shinohara, M., Gasior, S. L., Bishop, D. K. & Shinohara, A. Tid1/Rdh54 promotes colocalization of Rad51 and Dmc1 during meiotic recombination. *Proc. Natl Acad. Sci. USA* **97,** 10814–10819 (2000).
52. Petukhova, G., Stratton, S. & Sung, P. Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature* **393,** 91–94 (1998).
53. Petukhova, G., Sung, P. & Klein, H. Promotion of Rad51-dependent D-loop formation by yeast recombination factor Rdh54/Tid1. *Genes Dev.* **14,** 2206–2215 (2000).
54. Dresser, M. E. *et al.* DMC1 functions in a *Saccharomyces cerevisiae* meiotic pathway that is largely independent of the *RAD51* pathway. *Genetics* **147,** 533–544 (1997).
55. Catlett, M. G. & Forsburg, S. L. *Schizosaccharomyces pombe* Rdh54 (TID1) acts with Rhp54 (RAD54) to repair meiotic double-strand breaks. *Mol. Biol. Cell* **14,** 4707–4720 (2003).
56. Wesoly, J. *et al.* Differential contributions of mammalian Rad54 paralogs to recombination, DNA damage repair, and meiosis. *Mol. Cell. Biol.* **26,** 976–989 (2006).
57. Van Komen, S., Petukhova, G., Sigurdsson, S., Stratton, S. & Sung, P. Superhelicity-driven homologous DNA pairing by yeast recombination factors Rad51 and Rad54. *Mol. Cell* **6,** 563–572 (2000).
58. Ristic, D., Wyman, C., Paulusma, C. & Kanaar, R. The architecture of the human Rad54–DNA complex provides evidence for protein translocation along DNA. *Proc. Natl Acad. Sci. USA* **98,** 8454–8460 (2001).
59. Jaskelioff, M., Van Komen, S., Krebs, J. E., Sung, P. & Peterson, C. L. Rad54p is a chromatin remodeling enzyme required for heteroduplex DNA joint formation with chromatin. *J. Biol. Chem.* **278,** 9212–9218 (2003).
60. Alexeev, A., Mazin, A. & Kowalczykowski, S. C. Rad54 protein possesses chromatin-remodeling activity stimulated by the Rad51–ssDNA nucleoprotein filament. *Nature Struct. Biol.* **10,** 182–186 (2003).
61. Solinger, J. A., Kiianitsa, K. & Heyer, W. D. Rad54, a Swi2/Snf2-like recombinational repair protein, disassembles Rad51:dsDNA filaments. *Mol. Cell* **10,** 1175–1188 (2002).
62. Leu, J. Y., Chua, P. R. & Roeder, G. S. The meiosis-specific Hop2 protein of *S. cerevisiae* ensures synapsis between homologous chromosomes. *Cell* **94,** 375–386 (1998).
63. Gerton, J. L. & DeRisi, J. L. Mnd1p: an evolutionarily conserved protein required for meiotic recombination. *Proc. Natl Acad. Sci. USA* **99,** 6895–6900 (2002).
64. Petukhova, G. V., Romanienko, P. J. & Camerini-Otero, R. D. The Hop2 protein has a direct role in promoting interhomolog interactions during mouse meiosis. *Dev. Cell* **5,** 927–936 (2003).
65. Tsubouchi, H. & Roeder, G. S. The Mnd1 protein forms a complex with Hop2 to promote homologous chromosome pairing and meiotic double-strand break repair. *Mol. Cell. Biol.* **22,** 3078–3088 (2002).
66. Zierhut, C., Berlinger, M., Rupp, C., Shinohara, A. & Klein, F. Mnd1 is required for meiotic interhomolog repair. *Curr. Biol.* **14,** 752–762 (2004).
67. Henry, J. M. *et al.* Mnd1/Hop2 facilitates Dmc1-dependent interhomolog crossover formation in meiosis of budding yeast. *Mol. Cell. Biol.* **26,** 2913–2923 (2006).
68. Enomoto, R. *et al.* Stimulation of DNA strand exchange by the human TBPIP/Hop2–Mnd1 complex. *J. Biol. Chem.* **281,** 5575–5581 (2006).
69. Hayase, A. *et al.* A protein complex containing Mei5 and Sae3 promotes the assembly of the meiosis-specific RecA homolog Dmc1. *Cell* **119,** 927–940 (2004).
70. Tsubouchi, H. & Roeder, G. S. The budding yeast Mei5 and Sae3 proteins act together with Dmc1 during meiotic recombination. *Genetics* **168,** 1219–1230 (2004).
71. Akamatsu, Y., Dziadkowiec, D., Ikeguchi, M., Shinagawa, H. & Iwasaki, H. Two different Swi5-containing protein complexes are involved in mating-type switching and recombination repair in fission yeast. *Proc. Natl Acad. Sci. USA* **100,** 15770–15775 (2003).
72. Ellermeier, C., Schmidt, H. & Smith, G. R. Swi5 acts in meiotic DNA joint molecule formation in *Schizosaccharomyces pombe*. *Genetics* **168,** 1891–1898 (2004).
73. Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436,** 1053–1057 (2005).
74. Blat, Y., Protacio, R. U., Hunter, N. & Kleckner, N. Physical and functional interactions among basic chromosome organizational features govern early steps of meiotic chiasma formation. *Cell* **111,** 791–802 (2002).
75. Smith, A. V. & Roeder, G. S. The yeast Red1 protein localizes to the cores of meiotic chromosomes. *J. Cell Biol.* **136,** 957–967 (1997).
76. Klein, F. *et al.* A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. *Cell* **98,** 91–103 (1999).
77. Xu, H., Beasley, M. D., Warren, W. D., van der Horst, G. T. & McKay, M. J. Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis. *Dev. Cell* **8,** 949–961 (2005).

Acknowledgements Work from the authors' laboratory is supported by a grant from the US National Institutes of Health (to S.K.). M.J.N. is supported in part by a fellowship from the Human Frontiers Science Program. S.K. is a Leukemia and Lymphoma Society Scholar.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to S.K. (keeneys@mskcc.org).

© 2006 Nature Publishing Group

ARTICLES

A soft-bodied mollusc with radula from the Middle Cambrian Burgess Shale

Jean-Bernard Caron[1], Amélie Scheltema[2], Christoffer Schander[3] & David Rudkin[1]

***Odontogriphus omalus* was originally described as a problematic non-biomineralized lophophorate organism. Here we re-interpret *Odontogriphus* based on 189 new specimens including numerous exceptionally well preserved individuals from the Burgess Shale collections of the Royal Ontario Museum. This additional material provides compelling evidence that the feeding apparatus in *Odontogriphus* is a radula of molluscan architecture comprising two primary bipartite tooth rows attached to a radular membrane and showing replacement by posterior addition. Further characters supporting molluscan affinity include a broad foot bordered by numerous ctenidia located in a mantle groove and a stiffened cuticular dorsum. *Odontogriphus* has a radula similar to *Wiwaxia corrugata* but lacks a scleritome. We interpret these animals to be members of an early stem-group mollusc lineage that probably originated in the Neoproterozoic Ediacaran Period, providing support for the retention of a biomat-based grazing community from the late Precambrian Period until at least the Middle Cambrian.**

Burgess Shale-type deposits in Lower and Middle Cambrian strata yield a number of 'problematic' organisms that can potentially reveal key steps in the origin, relationship and evolution of phyla[1]. *Odontogriphus omalus* is one of the most enigmatic fossils from the Middle Cambrian Burgess Shale[2]. This animal, known originally from a single, incomplete and poorly preserved specimen, was described as a dorso-ventrally flattened and possibly annulated organism[2]. A conspicuous U-shaped feeding apparatus flanked by tooth-like structures was thought to be reminiscent of the lophophore of brachiopods, phoronids and ectoprocts, with a possible connection with some Cambrian conodonts[2]. The view that *Odontogriphus* could have had a key role in chordate evolution[3] has remained marginal, especially after the subsequent discovery of conodont animals bearing no resemblance to *Odontogriphus*[4]. Hypothetical tentacles originally reconstructed around the tooth-like structures are themselves dubious, and cannot be used to refer *Odontogriphus* to a known lophophorate group[4]. The suggestions that *Odontogriphus* could be related to the Early Cambrian problematic fossil *Vetustovermis* or to the poorly known Late Permian *Bowengriphus*[5] are very unlikely based on the abundant and exceptionally well preserved new material presented in this study. *Odontogriphus* shares a virtually identical radula with the non-calcified scleritome-bearing animal *Wiwaxia*[6], supporting the idea that both organisms are stem-group molluscs, contrary to views that *Wiwaxia* was a polychaete[7] or a stem-group polychaete[8] (but see ref. 9). Our study provides new insights into the origin and early evolution of the Mollusca, which together with the discovery of fossils showing probable molluscan affinities from the latest Precambrian (*Kimberella*[10]) and Early Cambrian strata (*Halkieria*[8,11]) confirm that the origin of eutrochozoans is deeply rooted and predates the Cambrian explosion[12].

Stem-group Mollusca
Odontogriphus omalus Conway Morris, 1976

Holotype. USNM 196169, (35K), National Museum of Natural History, Washington DC, USA.

Referred material and locality. One-hundred and eighty-eight specimens (Royal Ontario Museum) from the Greater Phyllopod Bed on Fossil Ridge (including one from talus) and one talus specimen from Mount Stephen (S7), Yoho National Park, British Columbia, Canada.

Horizon. Middle Cambrian Burgess Shale Formation.

Preservation. Specimens appear as black reflective films on a dark mudstone matrix and often co-occur with large sheets of the cyanobacterium *Morania* (Figs 1a, b and 2). The radula retains some original three-dimensionality, but its original composition has not been preserved (energy dispersive X-ray spectrometer (EDS) analyses show no difference between the radula and matrix). Evidence of decay is rare and most specimens are preserved parallel to bedding planes, implying that the body was flattened dorso-ventrally and that animals were buried very rapidly with limited or no transport (see ref. 13). *Odontogriphus* represents less than 0.5% of 50,900 individuals in the Greater Phyllopod Bed community[14].

Revised diagnosis. Bilaterally symmetrical oval body, parallel-sided, compressed dorso-ventrally. Anterior and posterior are semi-circular in outline and of similar size. The mouth is ventral with two primary bipartite tooth rows attached to a radular membrane and showing replacement by posterior addition. The gut is straight with a large stomach, narrow intestine and a sub-terminal anus. Simple ctenidia are present in a groove running laterally and posteriorly around a muscular foot. With non-biomineralized, stiffened cuticular mantle dorsum lacking sclerites.

Description. Specimens range from 3.3 to 125 mm in length (mean = 47.5 mm, s.d. = 29, $n = 89$) and from 1.5 to 43 mm in width (mean = 18.5 mm, s.d. = 10.5, $n = 123$) (Supplementary Fig. I). Length-to-width ratio is identical in juveniles and adults ($L = 2.74W$, $r_s = 0.97$, $P < 0.0001$, $n = 78$), demonstrating isometric growth. Symmetry of the body outline is maintained in all complete specimens, even those in which internal features have been preserved asymmetrically (Fig. 1h), implying that the dorsal body coverage is relatively stiff. The radula is located at about 15% of the total body length from the anterior margin on the mid-longitudinal

[1]Department of Natural History-Palaeobiology, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada. [2]Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [3]University of Bergen, Department of Biology, PO box 7800, N-5020 Bergen, Norway.

© 2006 Nature Publishing Group

axis (Fig. 1a, b, g, h, j, k). It typically consists of two rows of teeth preserved at different angles (Supplementary Fig. II) and is bipartite (that is, distichous) with paired mirror-image teeth connected axially (1:0:1) (Fig. 1c). There are usually seven pointed or rounded denticles on each tooth with the longest ones positioned laterally (Fig. 1c). In 81% of radulae the posterior row is as wide as or wider than the anterior row, the widest up to 5.3 mm; radula growth is proportional to body size. One or two faint, narrow posterior rows are visible in many individuals (Fig. 1d, e), implying, first, that new rows were periodically formed posteriorly, but only the anteriormost two rows were fully functional; and second, because the distance between the rows is closely similar, they were added at regular intervals. The anteriormost row was sloughed off periodically and occasionally ingested (Supplementary Fig. III), the sloughed off rows were replaced by new ones from behind. Isolated radulae with two rows of paired teeth (Fig. 1c) imply that the rows of teeth were connected by a strong decay-resistant material, a radular membrane (Fig. 1e).

A circular structure surrounding the radula is interpreted to represent the mouth and pharynx (Fig. 1f). The oesophagus is straight and narrow (Fig. 1f). It expands posteriorly into a straight stomach, which is usually as wide anteriorly as the radula itself (Fig. 1b, f–h, k). The stomach narrows posteriorly towards a straight intestine (Fig. 1b, g), and the anus is sub-terminal (Fig. 1g). Broad transverse wrinkles, parallel to each other and usually straight, sometimes occur across the midsection of the body (Fig. 1a, j). These vary in number between specimens and are only present in the central region of the body, demonstrating that they are not expressions of internal or surface segmentation. Wrinkles probably resulted from compression of a thickened central structure that we interpret to be ventral and thus probably represents a muscular sole. As a ventral sole it is limited anteriorly by the radula and thus lies posterior to the mouth. The sole is surrounded on all its sides except the front by darker and serially identical structures that we interpret to represent ctenidia (Fig. 1a, b, g–k). Ctenidia (up to 100) are sometimes separated from the mantle by a thin layer of sediment

Figure 1 | *Odontogriphus omalus* from the Middle Cambrian Burgess Shale. All specimens are preserved dorso-ventrally, anterior to the top. **a**, Royal Ontario Museum (ROM)57712, complete specimen showing the sole (crescentic wrinkles) lying on the surface of the cyanobacterium *Morania*. **b**, ROM57725, nearly complete specimen showing putative paired gonads or digestive glands and the cyanobacterium *Morania*. **c**, ROM57713, backscattered image of an isolated two-rowed radula. **d**, ROM57716, three-rowed radula. **e**, ROM57717, four-rowed radula with putative traces of the radular membrane. **f**, ROM57714, view of the mouth area and anterior end of the stomach. **g**, ROM57721, complete specimen showing the intestine and gut content. **h**, **i**, ROM57720 complete specimen. **h**, Overall view; **i**, detail of ctenidia from **h**. **j**, ROM57723, complete specimen showing the sole (wrinkles). **k**, ROM57724, complete specimen with paired salivary glands and ctenidia. Scale bars: **a**, 2 cm; **g**, **h**, **k**, 1 cm; **b**, **f**, **i**, **j**, 5 mm; **c**–**e**, 1 mm. an, anus; cr, crack; ct, ctenidia; ctg, ctenidia groove; dg?, digestive glands?; gc, gut content; go?, gonads?; in, intestine; ma, mantle; mo, *Morania*; oe, oesophagus; ph, pharynx and mouth area; ra, radula; ram?, radular membrane?; r1, r2, r3, r4, tooth rows; sg, salivary glands; so, sole; st, stomach.

© 2006 Nature Publishing Group

Figure 2 | Reconstruction of a colony of *Odontogriphus omalus* grazing on the cyanobacterium *Morania*. Illustration by Marianne Collins (copyright 2006).

(Fig. 1h, i), demonstrating that they are located in a narrow groove that we interpret to be a mantle cavity. A pair of circular structures interpreted as salivary glands flanks the radula (Fig. 1h, j, k). Another pair of structures, elongate and ovoid, composed of a bundle of fibrous elements, is preserved, one element on either side of the posterior part of the stomach and anterior part of the gut (Fig. 1b). Our interpretation is that they are gonads or digestive glands.

Discussion

The phylogenetic tree (Fig. 3) depicts the total-group Mollusca based on our interpretation of the morphology of *Odontogriphus omalus* (Fig. 2) and *Wiwaxia corrugata* (Fig. 4). It does not take into account many early mollusc-like forms of uncertain affinities that may represent various stem-group eutrochozoans (that is, Machaeridia, *Acaenoplax*, probivalvia, Hyolitha).

Odontogriphus and perhaps the Ediacaran form *Kimberella*[10] possess distinctive characters that place them in the molluscs before the acquisition of a calcified dorsum. That is, they or their closest ancestor arose in the latest Neoproterozoic (certain for *Kimberella*) following the node designating separation of the Annelida and Mollusca stems (Fig. 3, stem 1). The Mollusca ancestor is thought to have been a creeping and non-segmented bilateral animal[15] whereas the Annelida probably descended from a segmented parapodia-bearing form[16], possibly a sister group of the molluscan ancestor[15]. Mesodermal segmentation (metamerism) involving the coelom gave rise to the stem annelid (Fig. 3, stem 2). Non-coelomic iteration of organ systems, or seriation[15], led to the ancestral Mollusca (Fig. 3, stem 3); this condition is present in *Odontogriphus* as iterated ctenidia and radula teeth.

The characters held in common by *Odontogriphus* and *Kimberella*, the latter without preserved internal anatomy, are a dorso-ventrally flattened ovoid shape; large size; a cuticular dorsal exoskeleton shown by the integrity of the dorsum relative to soft anatomy (compare Fig. 1h with *Kimberella* (Fig. 1h of ref. 10)); a non-cuticularized ventral sole; and iterated structures (Fig. 3, stem 3). These are evolutionary novelties that arose since the estimated time of the last common ancestor of bilaterians[17,18]. If the interpretation of *Kimberella* as an early mollusc-like organism with radula is correct[19] these characteristics, combined with the soft and internal anatomy of *Odontogriphus*, define the plesiomorphic molluscan morphologies: pre-eminently and unique to Mollusca, an anterior radula with periodically sloughed off and replaced rows of teeth on a radular membrane[20] (certain for *Odontogriphus*). The acquisition of a ventral mantle groove with replicated ctenidia represents a younger evolutionary innovation (Fig. 3, stem 4).

Wiwaxia corrugata with its dorsum covered by non-calcified sclerites co-occurs with *Odontogriphus* in the Burgess Shale and has a nearly identical radula (Fig. 4) with a tooth morphology similar to plesiomorphic Neomeniomorpha[20]. Its affinities with the polychaetes and annelids have been refuted based on absence of externally expressed mesodermal segmentation, prostomial appendages and parapodia[9]. On the basis of its radula, ovoid shape, lack of segmentation, and zonation of different sclerite types similar to that in *Halkieria*[8,11], it is considered here to belong to the clade Mollusca (Fig. 3, stem 5).

Initial acquisition of a calcified exoskeleton and subsequent rapid radiation of shelly forms in the Early Cambrian[21] marks the turning point for molluscan diversity. The process began with nucleation of sclerites within the epidermis, which then pushed through the overlying cuticle (Fig. 3, stem 6), a process retained by Polyplacophora and Aplacophora (Neomeniomorpha and Chaetodermomorpha[22]), and

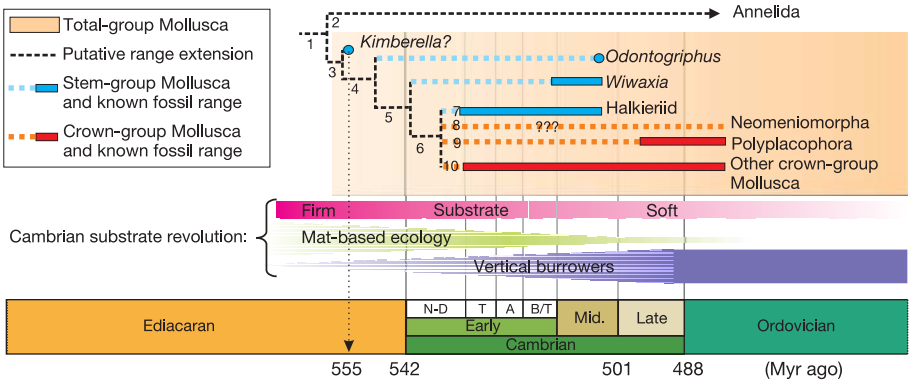


Figure 3 | Evolutionary tree of the molluscs in the context of the Neoproterozoic–Cambrian substrate revolution[35]. 1: Protostome bilaterian; serial replication; triploblastic. 2: Segmentation by coelomic metameres. 3: Large size; with iteration but not coelomic segmentation; ovoid; dorso-ventrally flattened; stiffened cuticular dorsum; flat, non-cuticularized ventral sole; radula of iterated, paired mirror-image teeth and radular membrane (certain for *Odontogriphus*); feeding on biomat? 4: Groove (mantle cavity) between dorsum and ventrum with serial ctenidia; paired salivary glands; straight digestive tract; nervous system ladder-like?; coelom posterior, restricted to reproductive and excretory organs? 5: Non-calcified scleritome, sclerites arranged in three mirror-image longitudinal zones. 6: Calcification of epidermally nucleated sclerites that pass through cuticle; calcified shell from serial shell fields; no periostracum from periostracal groove of mantle lobe. 7: Two shell fields. 8: Tubiform; reduced foot; sclerites in 1–3 longitudinal rows beside foot groove; progenetic loss of gills and shells; embryological evidence of vestigial shell fields. 9: Eight or more shell fields; sclerites not in longitudinal zones. 10: Loss of sclerites and serial shell fields; true periostracum secreted from mantle lobe; shells paired or single; reduction of gills; further variety of body plans. A, Atdabanian; B/T, Botomian/Toyonian; N-D, Nemakit-Daldynian; T, Tommotian.

© 2006 Nature Publishing Group

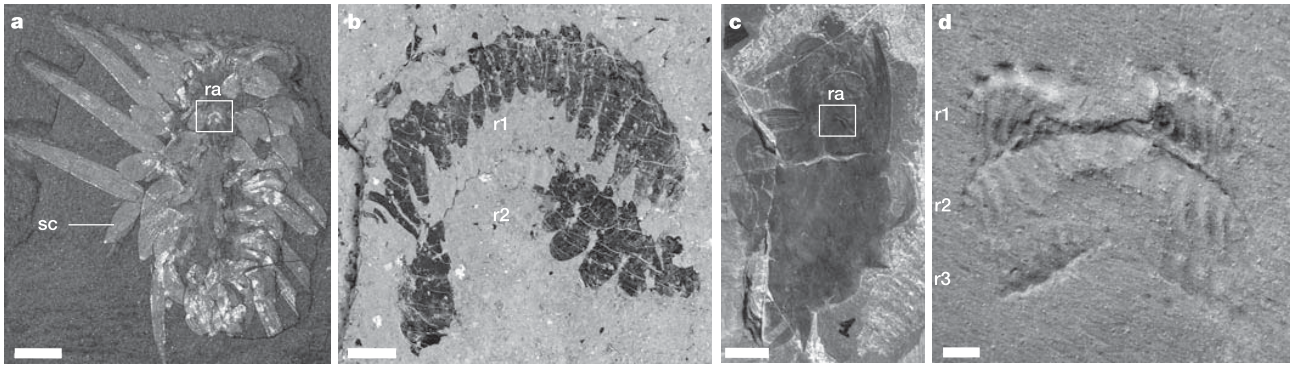


Figure 4 | *Wiwaxia corrugata* from the Middle Cambrian Burgess Shale. All specimens are preserved ventrally, anterior to the top. **a**, **b**, ROM57707, pictures courtesy N. Butterfield. **a**, Complete specimen and location of the radula. **b**, Backscattered image of the radula. **c**, **d**, ROM57726. **c**, Complete specimen and location of the radula. **d**, Detail of the radula highlighted with ammonium chloride sublimate. Note the presence of three rows of teeth. Scale bars: **a**, **c**, 5 mm; **b**, **d**, 250 μm. ra, radula; r1, r2, r3, tooth rows; sc, sclerite.

may have been the same for *Halkieria*. Shells in the Polyplacophora and possibly *Halkieria* have been thought to be the result of merging of sclerites[23]. However, it is more likely that the acquisition of shell fields provided the functional ability to deposit calcareous shell (developmentally separate from sclerite formation in the Polyplacophora[22,24]). The polytomy of *Halkieria* (and other halkieriids), Neomeniomorpha and Polyplacophora (Fig. 3, stems 7, 8 and 9) includes on the polyplacophoran stem the extinct *Matthevia*[25] and the Multiplacophora[26]. *Halkieria*, originally described as probably a molluscan-grade organism, in part due to the presence of a putative radula[11] (but see ref. 8), has been interpreted most recently as a sister taxon to the Polyplacophora[27]. However, affinities of *Halkieria* still remain highly contentious. Neomeniomorpha have several morphologies homologous to those in Polyplacophora[28]. Some of these are plesiomorphies of the Mollusca (Fig. 3, stem 3), but others cannot be accommodated on the tree as plesiomorphies, including eight shell fields, embryologically vestigial in a neomeniomorph[29], manner of sclerite deposition[22] and unique epidermal gland cells[30,31]. The molecular evidence for the relationship among the Polyplacophora and the two taxa of Aplacophora is still ambiguous.

The evidence from *Odontogriphus*, *Wiwaxia*, *Halkieria* and *Kimberella* indicates that early molluscs were not small. Diversity of body form increased with the loss of sclerites and the advent of a true periostracum secreted from a groove in the mantle lobe, as found in all extant Mollusca except the Aplacophora and Polyplacophora[22] (Fig. 3, stem 10). This diversification was concomitant with the appearance of crown-group (except Scaphopoda) and other stem-group molluscs by the end of the Cambrian[32].

The Cambrian substrate revolution and ecological implications

Body (*Kimberella*) and trace (*Radulichnus*) fossil evidence from Ediacaran shallow marine sediments demonstrates the establishment of a bilaterian microphagous mat-grazing guild by at least 555 million years ago[19,33]. The subsequent transition from Neoproterozoic biomat-dominated sea floors to Phanerozoic-style seafloor conditions, characterized by increasingly fluidized substrates, was driven by a shift to more intensive and vertically oriented bioturbation across the Precambrian/Cambrian boundary[34] (see also Fig. 2 of ref. 35). The redistribution of extensive microbial mats and biofilms (along with metazoan grazers) from the open shallow marine to stressed nearshore and shelf-edge to deep-sea settings[35] was not abrupt, but took place over a protracted interval in the Cambrian during which relict mat-based communities persisted in at least some marine environments[36], including the Burgess Shale[34]. This persistence is evident from the intimate association of *Odontogriphus* and *Wiwaxia* with dense, sheet-like aggregates of the fossil cyanobacterium *Morania*[37], which often cover extensive bedding surfaces[14] (Figs 1a and 2). *Morania* probably provided a food source and stable substrate for an array of Middle Cambrian benthic grazers adapted to Neoproterozoic-style substrates[38].

Odontogriphus joins a handful of Cambrian fossils that probably represent surviving Neoproterozoic lineages (for example, see refs 39–41). Explanations for the disappearance of other soft-bodied elements of the Ediacaran biota ('vendobionts') vary from a mass extinction following global environmental changes, to changes in taphonomic conditions, to the emergence of predators[42]. The widespread appearance of biomineralized organisms near the beginning of the Cambrian is believed to be in large part the consequence of rapidly expanding predatory selection pressure (for example, see ref. 43). However, the presence of a diverse soft-bodied biota in Cambrian Lagerstätten, including representatives of relict lineages, demonstrates that other survival strategies were in play. *Odontogriphus* would have been a prime target for predators, but it is possible that its featureless dorsum (Fig. 2) afforded a cryptic lifestyle on the substrate. In the case of the *Kimberella*–*Odontogriphus* lineage, reduction of suitable habitat could have been at least as important as direct predation pressure in explaining the apparent demise of the group after the Middle Cambrian.

METHODS

Digital photography used polarizing filters at the camera and light source to increase contrast of internal organs and other anatomical features. Backscattered electron images (BSE) were taken of radulae with a SEMCO Nanolab 7 at the Royal Ontario Museum with an acceleration voltage of 30 kV (15 kV for EDS analyses) (see ref. 44 for a successful application of this technique). Specimens were wrapped in aluminium foil to limit charging in a manner described previously[45].

Received 15 February; accepted 8 May 2006.

1. Budd, G. E. The Cambrian fossil record and the origin of the phyla. *Integr. Comp. Biol.* **43**, 157–165 (2003).
2. Conway Morris, S. A new Cambrian lophophorate from the Burgess Shale of British Columbia. *Palaeontology* **19**, 199–222 (1976).
3. Dzik, J. *Yunnanozoon* and the ancestry of chordates. *Acta Palaeont. Pol.* **40**, 341–360 (1995).
4. Briggs, D. E. G. & Conway Morris, S. in *Problematic Fossil Taxa* (eds Hoffman, A. & Nitecki, M. H.) 167–183 (Oxford Univ. Press/Clarendon Press, New York, 1986).
5. Ritchie, A. & Edgecombe, G. D. An Odontogriphid from the Upper Permian of Australia. *Palaeontology* **44**, 861–874 (2001).
6. Conway Morris, S. The Middle Cambrian metazoan *Wiwaxia corrugata* (Matthew) from the Burgess Shale and *Ogygopsis* Shale, British Columbia, Canada. *Phil. Trans. R. Soc. Lond. B* **307**, 507–582 (1985).
7. Butterfield, N. J. A reassessment of the enigmatic Burgess Shale fossil *Wiwaxia corrugata* (Matthew) and its relationship to the polychaete *Canadia spinosa* Walcott. *Paleobiology* **16**, 287–303 (1990).
8. Conway Morris, S. & Peel, J. S. Articulated halkieriids from the Lower Cambrian of North Greenland and their role in early protostome evolution. *Phil. Trans. R. Soc. Lond. B* **347**, 305–358 (1995).

© 2006 Nature Publishing Group

9. Eibye-Jacobsen, D. A reevaluation of *Wiwaxia* and the polychaetes of the Burgess Shale. *Lethaia* **37,** 317–335 (2004).
10. Fedonkin, M. A. & Waggoner, B. M. The Late Precambrian fossil *Kimberella* is a mollusc-like bilaterian organism. *Nature* **388,** 868–871 (1997).
11. Conway Morris, S. & Peel, J. S. Articulated halkieriids from the Lower Cambrian of north Greenland. *Nature* **345,** 802–805 (1990).
12. Erwin, D. H. & Davidson, E. H. The last common bilaterian ancestor. *Development* **129,** 3021–3032 (2002).
13. Caron, J. B. & Jackson, D. A. Taphonomy of the Greater Phyllopod Bed Community, Burgess Shale. *Palaios* **21**(5), (in the press).
14. Caron, J. B. *Taphonomy and Community Analysis of the Middle Cambrian Greater Phyllopod Bed, Burgess Shale*. Thesis, Univ. Toronto (2005).
15. Valentine, J. W. *On the Origin of Phyla* (Univ. Chicago Press, Chicago/London, 2004).
16. Balavoine, G. & Adoutte, A. The segmented Urbilateria: A testable scenario. *Integr. Comp. Biol.* **43,** 137–147 (2003).
17. Peterson, K. J. *et al.* Estimating metazoan divergence times with a molecular clock. *Proc. Natl Acad. Sci. USA* **101,** 6536–6541 (2004).
18. Douzery, E. J. P., Snell, E. A., Bapteste, E., Delsuc, F. & Philippe, H. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl Acad. Sci. USA* **101,** 15386–15391 (2004).
19. Seilacher, A. Biomat-related lifestyles in the Precambrian. *Palaios* **14,** 86–93 (1999).
20. Scheltema, A. H., Kerth, K. & Kuzirian, A. M. Original molluscan radula: Comparisons among Aplacophora, Polyplacophora, Gastropoda, and the Cambrian fossil *Wiwaxia corrugata*. *J. Morphol.* **257,** 219–244 (2003).
21. Runnegar, B. in *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.) 77–87 (Oxford Univ. Press, New York, 1996).
22. Haas, W. Evolution of calcareous hardparts in primitive molluscs. *Malacologia* **21,** 403–418 (1981).
23. Bengtson, S. The cap-shaped Cambrian fossil *Maikhanella* and the relationship between coeloscleritophorans and molluscs. *Lethaia* **25,** 401–420 (1992).
24. Henry, J. Q., Okusu, A. & Martindale, M. Q. The cell lineage of the polyplacophoran, *Chaetopleura apiculata*: variation in the spiralian program and implications for molluscan evolution. *Dev. Biol.* **272,** 145–160 (2004).
25. Runnegar, B., Pojeta, J. Jr, Taylor, M. E. & Collins, D. New species of the Cambrian and Ordovician chitons *Matthevia* and *Chelodes* from Wisconsin and Queensland; evidence for the early history of polyplacophoran mollusks. *J. Paleontol.* **53,** 1374–1394 (1979).
26. Vendrasco, M. J., Wood, T. E. & Runnegar, B. N. Articulated Paleozoic fossil with 17 plates greatly expands disparity of early chitons. *Nature* **429,** 288–291 (2004).
27. Vinther, J. & Nielsen, C. The Early Cambrian *Halkieria* is a mollusc. *Zool. Scr.* **34,** 81–89 (2005).
28. Scheltema, A. H. Aplacophora as progenetic aculiferans and the coelomate origin of mollusks as the sister taxon of Sipuncula. *Biol. Bull.* **184,** 57–78 (1993).
29. Scheltema, A. H. & Ivanov, D. L. An aplacophoran postlarva with iterated dorsal groups of spicules and skeletal similarities to Paleozoic fossils. *Invertebr. Biol.* **21,** 1–10 (2002).
30. Scheltema, A. H., Tscherkassky, M. & Kuzirian, A. M. in *Microscopic Anatomy of Invertebrates* (eds Harrison, F. W. & Kohn, A. J.) 13–54 (Wiley-Liss, New York, 1994).
31. Fischer, F. P. Die Mantelpapillen und Stacheln von *Acanthochiton fascicularis* L. (Mollusca, Polyplacophora). *Zoomorphologie* **94,** 121–131 (1980).
32. Budd, G. E. & Jensen, S. A critical reappraisal of the fossil record of the bilaterian phyla. *Biol. Rev.* **75,** 253–295 (2000).
33. Seilacher, A., Buatois, L. A. & Mángano, L. G. Trace fossils in the Ediacaran-Cambrian transition: Behavioral diversification, ecological turnover and environmental shift. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **227,** 323–356 (2005).
34. Seilacher, A. & Pflüger, F. in *Biostabilization of Sediments* (eds Krumbein, W. E., Peterson, D. M. & Stal, L. J.) 97–105 (Bibliotheks und Informationssystem der Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany, 1994).
35. Bottjer, D. J., Hagadorn, J. W. & Dornbos, S. Q. The Cambrian substrate revolution. *GSA Today* **10,** 1–7 (2000).
36. Dornbos, S., Bottjer, D. & Chen, J.-Y. Evidence for seafloor microbial mats and associated metazoan lifestyles in Lower Cambrian phosphorites of Southwest China. *Lethaia* **37,** 127–137 (2004).
37. Satterthwait, D. F. *Paleobiology and Paleoecology of Middle Cambrian Algae from Western North America*. Thesis, Univ. California (1976).
38. Dornbos, S. Q., Bottjer, D. J. & Chen, J.-Y. Paleoecology of benthic metazoans in the Early Cambrian Maotianshan Shale biota and the Middle Cambrian Burgess Shale biota: evidence for the Cambrian substrate revolution. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **220,** 47–67 (2005).
39. Conway Morris, S. Ediacaran-like fossils in Cambrian Burgess Shale-type faunas of North America. *Palaeontology* **36,** 593–635 (1993).
40. Hagadorn, J. W., Fedo, C. M. & Waggoner, B. M. Early Cambrian Ediacaran-type fossils from California. *J. Paleontol.* **74,** 731–740 (2000).
41. Jensen, S., Gehling, J. G. & Droser, M. L. Ediacara-type fossils in Cambrian sediments. *Nature* **393,** 567–569 (1998).
42. Narbonne, G. M. The Ediacara biota: Neoproterozoic origin of animals and their ecosystems. *Annu. Rev. Earth Planet. Sci.* **33,** 421–442 (2005).
43. Bengtson, S. in *The Fossil Record of Predation* (eds Kowalewski, M. & Kelley, P. H.) 289–317 (The Paleontological Society, New Haven, 2002).
44. Orr, P. J., Kearns, S. L. & Briggs, D. E. G. Backscattered electron imaging of fossils exceptionally-preserved as organic compressions. *Palaios* **17,** 110–117 (2002).
45. Allison, P. A. Phosphatized soft-bodied squids from the Jurassic Oxford Clay. *Lethaia* **21,** 403–410 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Bengtson, G. Budd, S. Conway Morris, S. Dornbos, W. Hagadorn, T. Høisaeter and B. Runnegar for reviewing different drafts of this paper. N. Butterfield suggested using the BSE technique and provided unpublished pictures of *Wiwaxia*. Our research was in part supported by a Post-Doctoral Natural Sciences and Engineering Research Council of Canada grant (to J.-B.C.) and by a Swedish Research Council grant (to C.S.). Permission to collect Burgess Shale specimens was given by Parks Canada to D. Collins. We are thankful to M. Back and J. Waddington from the Royal Ontario Museum for technical help. This is Royal Ontario Museum Burgess Shale Research Project number 6.

Author contributions All four authors made significant contributions to this article.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.-B.C. (jcaron@rom.on.ca).

© 2006 Nature Publishing Group

ARTICLES

Neuronal ensemble control of prosthetic devices by a human with tetraplegia

Leigh R. Hochberg[1,2,4], Mijail D. Serruya[2,3], Gerhard M. Friehs[5,6], Jon A. Mukand[7,8], Maryam Saleh[9]†, Abraham H. Caplan[9], Almut Branner[10], David Chen[11], Richard D. Penn[12] & John P. Donoghue[2,9]

Neuromotor prostheses (NMPs) aim to replace or restore lost motor functions in paralysed humans by routeing movement-related signals from the brain, around damaged parts of the nervous system, to external effectors. To translate preclinical results from intact animals to a clinically useful NMP, movement signals must persist in cortex after spinal cord injury and be engaged by movement intent when sensory inputs and limb movement are long absent. Furthermore, NMPs would require that intention-driven neuronal activity be converted into a control signal that enables useful tasks. Here we show initial results for a tetraplegic human (MN) using a pilot NMP. Neuronal ensemble activity recorded through a 96-microelectrode array implanted in primary motor cortex demonstrated that intended hand motion modulates cortical spiking patterns three years after spinal cord injury. Decoders were created, providing a 'neural cursor' with which MN opened simulated e-mail and operated devices such as a television, even while conversing. Furthermore, MN used neural control to open and close a prosthetic hand, and perform rudimentary actions with a multi-jointed robotic arm. These early results suggest that NMPs based upon intracortical neuronal ensemble spiking activity could provide a valuable new neurotechnology to restore independence for humans with paralysis.

Hundreds of thousands of people suffer from forms of motor impairment in which intact movement-related areas of the brain cannot generate movements because of damage to the spinal cord, nerves, or muscles[1]. Paralysing disorders profoundly limit independence, mobility and communication. Current assistive technologies rely on devices for which an extant function provides a signal that substitutes for missing actions. For example, cameras can monitor eye movements that can be used to point a computer cursor[2]. Although these surrogate devices have been available for some time, they are typically limited in utility, cumbersome to maintain, and disruptive of natural actions. For instance, gaze towards objects of interest disrupts eye-based control. By contrast, an NMP is a type of brain–computer interface (BCI) that can guide movement by harnessing the existing neural substrate for that action—that is, neuronal activity patterns in motor areas. An ideal NMP would provide a safe, unobtrusive and reliable signal from the disconnected motor area that could restore lost function. Neurons in the primary motor cortex (MI) arm area of monkeys, for example, provide information about intended arm reaching trajectories[3–5], but this command signal would work for an NMP only if neural signals persist and could be engaged by intention in paralysed humans.

In concept, NMPs require a sensor to detect the activity of multiple neurons, a decoder to translate ensemble firing patterns into motor commands, and, typically, a computer gateway to engage effectors. BrainGate (Cyberkinetics, Inc.) is an NMP system under development and in pilot trials in people with tetraparesis from spinal cord injury, brainstem stroke, muscular dystrophy, or amyotrophic lateral sclerosis. Currently, this system consists of a chronically implanted sensor and external signal processors developed from preclinical animal studies (see Methods)[6–8]. The participant described in this report, the first in the BrainGate trial, is a 25-yr-old male (MN) who sustained a knife wound in 2001 that transected the spinal cord between cervical vertebrae C3–C4, resulting in complete tetraplegia (C4 ASIA A)[9]. The array was implanted in June 2004 into the MI arm area 'knob'[10], as identified on pre-operative magnetic resonance imaging (MRI) (Fig. 1c). Post-operative recovery was uneventful. The data presented here are derived from 57 consecutive recording sessions from 14 July 2004 to 12 April 2005 (9 months).

Signal quality and variety

Action potentials were readily observable on multiple electrodes, indicating that MI neural spiking persists three years after SCI, as suggested indirectly by functional MRI data[11–14]. Recorded signals ranged from qualitatively well-isolated single neurons to mixtures of a few different waveforms (Fig. 2a). Different waveform shapes were identified visually, using standard time-amplitude windows, but there was no further attempt to distinguish between well isolated and intermixed waveforms, both of which we refer to in this report as 'units'. An average of 26.9 ± 14.2 units were observed each day (range 3–57), with mean peak-to-peak spike amplitudes of 76.4 ± 25.0 μV (mean ± s.d., $n = 56$ sessions) (see Supplementary

[1]Department of Neurology, Massachusetts General Hospital, Brigham and Women's Hospital, and Spaulding Rehabilitation Hospital, Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114, USA. [2]Department of Neuroscience and Brain Science Program, and [3]Department of Engineering, Brown University, PO Box 1953, Providence, Rhode Island 02912, USA. [4]Center for Restorative and Regenerative Medicine, Rehabilitation Research and Development Service, Department of Veterans Affairs, Veterans Health Administration, 830 Chalkstone Avenue, Providence, Rhode Island 02908, USA. [5]Department of Clinical Neurosciences (Neurosurgery), Brown University, and [6]Department of Neurosurgery, Rhode Island Hospital, 120 Dudley Street, Suite 103, Providence, Rhode Island 02905, USA. [7]Department of Rehabilitation Medicine, Brown University, 593 Eddy Street, Providence, Rhode Island 02903, USA. [8]Sargent Rehabilitation Center, 800 Quaker Lane, Warwick, Rhode Island 02818, USA. [9]Cyberkinetics Neurotechnology Systems, Inc., 100 Foxborough Boulevard–Suite 240, Foxborough, Massachusetts 02035, USA. [10]Cyberkinetics Neurotechnology Systems, Inc., 391 Chipeta Way, Suite G, Salt Lake City, Utah 84108, USA. [11]Department of Physical Medicine and Rehabilitation, Rehabilitation Institute of Chicago, 345 E. Superior Street, 1146, Chicago, Illinois 60611, USA. [12]Department of Neurosurgery, University of Chicago Hospitals, 5841 S. Maryland Avenue, MC3026, Chicago, Illinois 60637, USA. †Present address: Graduate Program in Computational Neuroscience, University of Chicago, Chicago, Illinois 60637, USA.

© 2006 Nature Publishing Group

Information)[8]. Although details of local field potentials (LFPs) will be described in a subsequent report, we note that LFPs, which could be recorded simultaneously with spikes, resembled those observed in intact monkeys (Fig. 2b)[15]. A notable decrease in the number of recorded units was seen approximately 6.5 months after implantation and thereafter. Upon approval of a clinical protocol change at 10 months, which permitted impedance measurements, we observed a low impedance in 54 of the electrodes, consistent with a physical short circuit to ground in the array, cable, or connector, but not consistent with a biological event (for example, gliosis). The precise reason for this physical change remains under investigation.

Since this report was written, we added a second trial participant to the study, a 55-yr-old man who has had complete spinal cord injury at C4 since 1999. Recordings were collected starting in the seventh month after implant, after making an electrical contacts repair in the pedestal connector. We recorded an average of 53.2 ± 6.3 units per session during trial months 7–10, again demonstrating the presence of neural activity lasting many months. Another technical issue causing abrupt signal loss at most electrodes, which may be related to the original repair, occurred at month 11 in participant 2; the reason for this change is being evaluated.

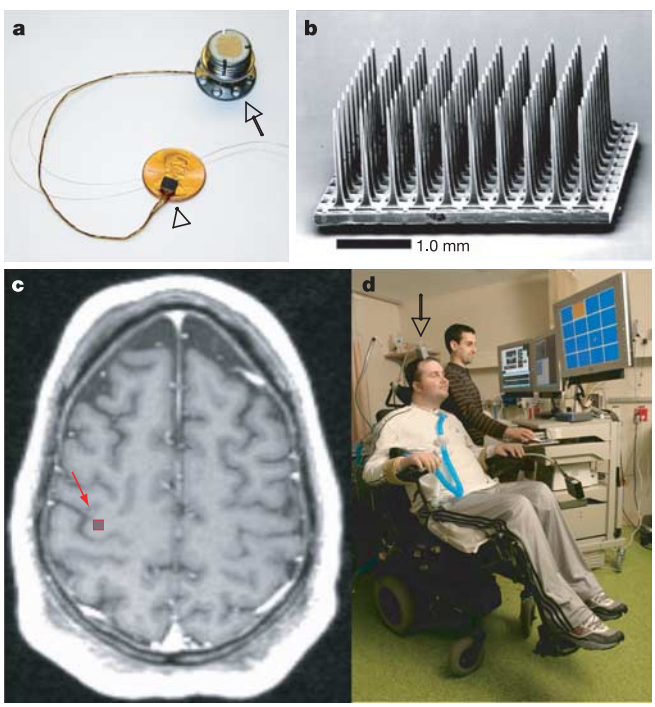


Figure 1 | Intracortical sensor and placement, participant 1. a, The BrainGate sensor (arrowhead), resting on a US penny, connected by a 13-cm ribbon cable to the percutaneous Ti pedestal (arrow), which is secured to the skull. Neural signals are recorded while the pedestal is connected to the remainder of the BrainGate system (seen in **d**). **b**, Scanning electron micrograph of the 100-electrode sensor, 96 of which are available for neural recording. Individual electrodes are 1-mm long and spaced 400 μm apart, in a 10 × 10 grid. **c**, Pre-operative axial T1-weighted MRI of the brain of participant 1. The arm/hand 'knob' of the right precentral gyrus (red arrow) corresponds to the approximate location of the sensor implant site. A scaled projection of the 4 × 4-mm array onto the precentral knob is outlined in red. **d**, The first participant in the BrainGate trial (MN). He is sitting in a wheelchair, mechanically ventilated through a tracheostomy. The grey box (arrow) connected to the percutaneous pedestal contains amplifier and signal conditioning hardware; cabling brings the amplified neural signals to computers sitting beside the participant. He is looking at the monitor, directing the neural cursor towards the orange square in this 16-target 'grid' task. A technician appears (A.H.C.) behind the participant.

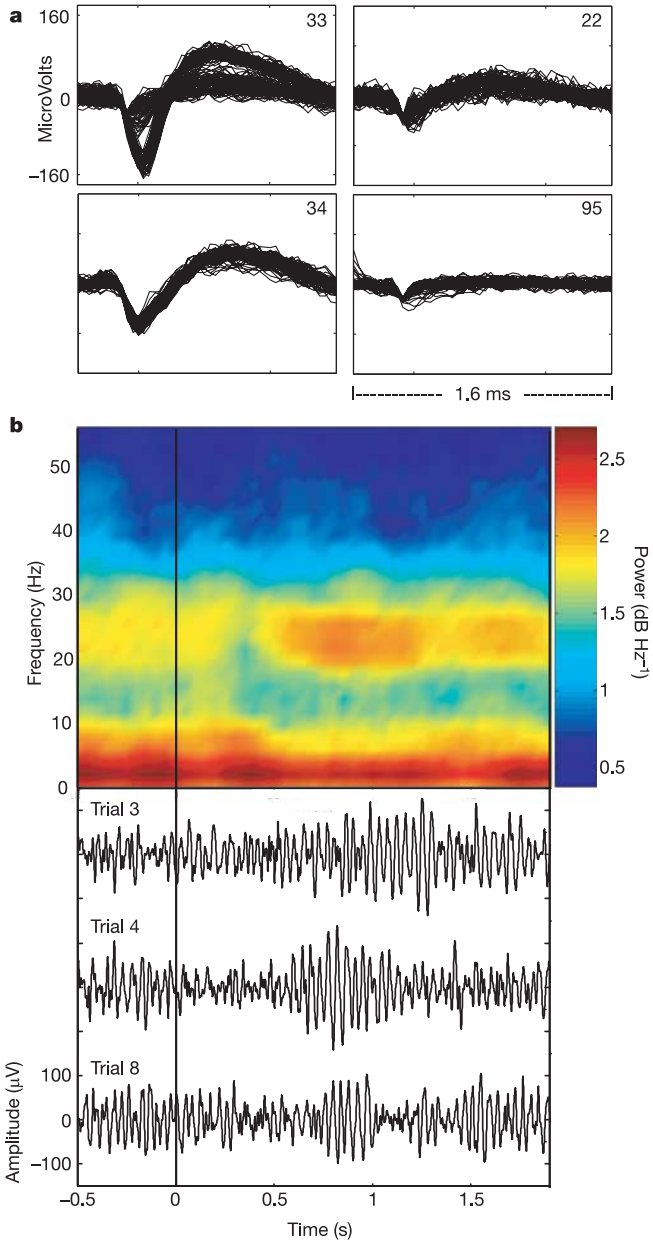


Figure 2 | Electrical recordings from a sample of four electrodes. a, Discriminated neural activity at electrodes 33, 34, 22, 95 ($n = 80$ superimposed action potentials for each unit). On electrode 33, two neuronal units could be reliably discriminated with peak to peak amplitudes of 206 and 56 μv, respectively. For electrode 34, a single unit is displayed. Electrode 22 illustrates a low-amplitude discriminated signal. Electrode 95 shows triggered noise. Data are from trial day 90 (90 days after array placement). **b**, Local field potentials during neural cursor control. In the bottom panel, three traces of electrical recording (bandpass: 10–100 Hz) from one electrode are shown 0.5 s before and 1.9 s after the go cue instructing MN to move the cursor from the centre position to a target at the top of the screen. In the top panel, a Thomson multi-taper time frequency analysis on each trial data segment was performed. This was done by sliding a 0.3-s window every 0.05 s, using a spectral resolution of 10 Hz. These power spectrograms were averaged across 20 trials to create the resulting pseudocolour power spectral density (PSD) plot. The diagram is aligned such that each point in the PSD plot corresponds to a time window 150 ms before and after an LFP. In the 20–30-Hz band, a decrease in power is seen approximately 300 ms after the go cue, followed by an increase in power from 550–1,200 ms after the go cue, which can also be appreciated in the raw, single trial data below.

© 2006 Nature Publishing Group

Modulation by intent

Imagined limb motions modulated neural firing on multiple electrodes, upon request, beginning at the first experimental session. Modulation was evaluated during four consecutive sessions when MN was asked to imagine a series of movements. This series revealed a rich variety of firing modulations largely consistent with patterns observed in monkey MI[16]. Importantly, this activity was evoked by imagined actions in this participant with cervical spinal cord injury. Figure 3a illustrates how certain neurons are selective for one imagined action (hands together/apart), whereas others recorded simultaneously are engaged by different imagined actions (elbow or wrist). This diversity includes neurons that fired with imagined hand or distal arm actions (for example, hand open/close, Fig. 3c) and those that fired during shoulder movements that were actually performed (see also Supplementary Fig. 1). Non-selective neurons, active with the onset of any imagined upper extremity action (Fig. 3b), were also observed. As shown in Fig. 3c and Supplementary Fig. 1, neurons fired in a relatively time-locked manner upon the request to imagine action. These results demonstrate a rich heterogeneity of firing patterns within a limited sample from a small MI region. This diversity is useful in creating a flexible control signal.

Linear filter construction

Use of MI neuronal ensemble activity as a control signal by persons with paralysis requires a novel approach to establishing a transform (filter function) between firing patterns and intended action. For each session, units were used to create a filter (see Methods) to provide a two-dimensional output signal displayed as cursor position on a monitor. A simple linear filter algorithm, identical to that used in intact monkeys, was used to create filters[17]. Unlike most studies performed using intact monkeys, where hand motions are known and kinematics are measured directly, we predicted MN's intended hand movements on the basis of

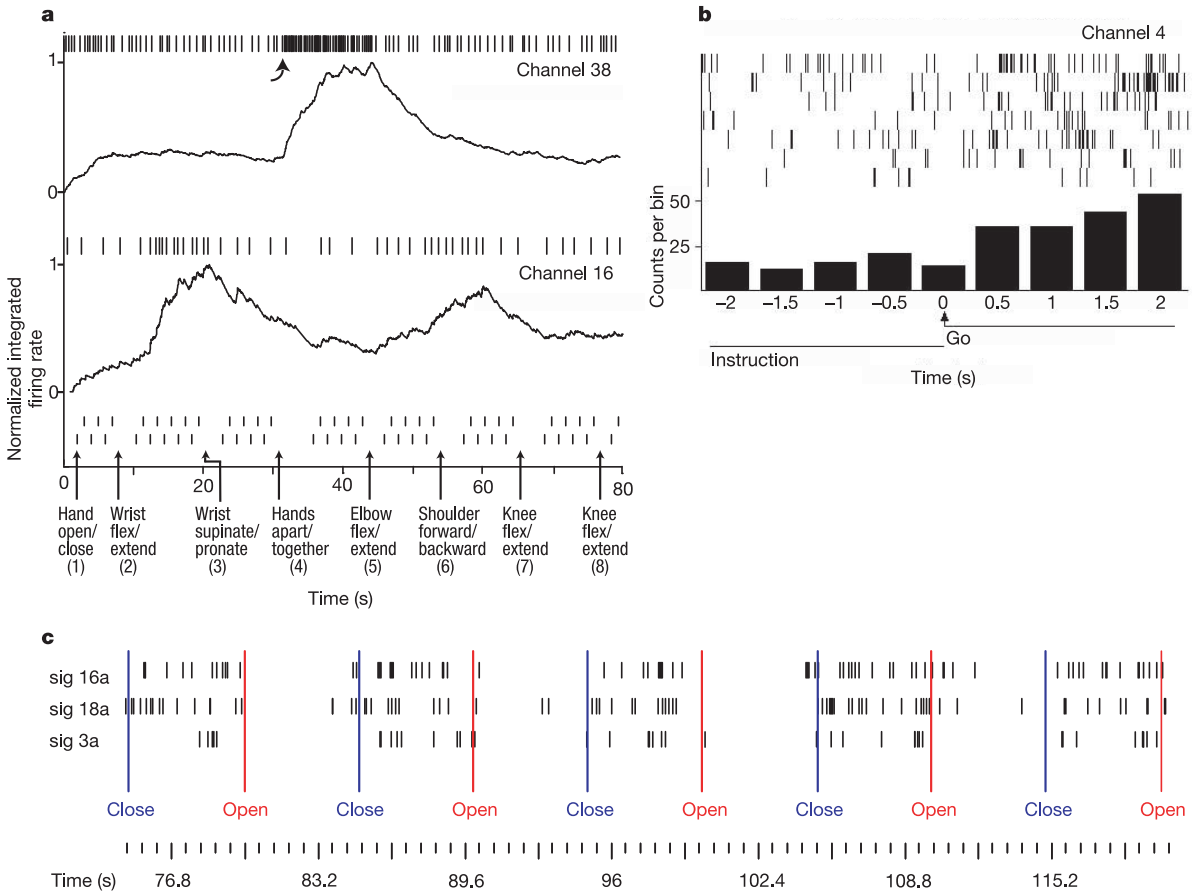


Figure 3 | Neuronal selectivity for imagined and performed movements. **a**, Over an 80-s period, MN was asked to imagine performing a series of left limb movements (which are described on the abscissa). Movement instruction time is indicated by a vertical arrow; the go cue for each alternating movement, presented as text on the video monitor, is indicated by a small vertical hash mark. Spiking activity of two simultaneously recorded units is displayed. Rasters indicate the time of each spike (thinned for visual clarity; every third spike is shown). Normalized, integrated firing rates (R) appear beneath each raster, as derived by the equation $R = [(R_{-1} + n)(1 - e^{-b/\tau})]$, where R_{-1} is the previous bin's integrated firing rate value, n = the number of spikes in the current bin, b = bin width, and τ = time constant; bin width = 50-ms window, time constant = 10 ms (adapted from ref. 28); normalization is achieved by dividing by the maximum integrated firing rate from each unit's spike train over the time period displayed. The top unit (channel 38) increases its firing rate (curved arrow) with the instruction to move both hands apart/together. The bottom, simultaneously recorded unit (channel 16) is activated most clearly after the instruction to flex/extend the wrist and to flex/extend the elbow or move the shoulder anteriorly and posteriorly. All movements are imagined except for shoulder movement, which MN actually performed. **b**, Go-cue-related activity modulation for a neuron recorded simultaneously with those in **a**. Each raster line is centred about the go cue, which requests that the patient imagine a movement; the seven raster lines represent the epochs surrounding each of the seven different movements in sequence **a**. The histogram displays the total number of spikes seen in each 500-ms bin. This neuron increased its firing rate during most imagined movement epochs, but demonstrated poor instruction selectivity compared to the neurons presented in **a**. Data are from day 161. **c**, Hand-instruction-related modulation for three simultaneously recorded neurons. MN was cued to open and close his hand by text instruction, presented on the screen. Go cues are indicated. Each vertical tick represents one action potential (spike). An increase in these neurons' firing rates is directly indicative of the intention to close the hand. Data are from day 90.

© 2006 Nature Publishing Group

instructed actions (instruction-based algorithms have also been reported in one study of intact monkeys[18]). Thus, for filter building, MN was asked to imagine manually tracking a 'technician's cursor' that was actually being moved by a technician-operated mouse through a succession of randomly placed visual targets (see Methods). The filter function was used to decode activity and drive a 'neural cursor'.

MI activity during neural cursor control

Features of neurons during neural cursor control resembled those expected from MI. Neurons in MI of intact monkeys characteristically begin to modulate their firing before movement onset and activity is tuned to hand movement direction[19–21]. To compare this neural activity with MI of a human with spinal cord injury, MN performed a step-tracking, 'centre-out' task using the neural cursor. The task requires that the neural cursor be moved from a centre target to one of four radially displaced targets (screen location: up, down, left, right; see Supplementary Video 1). For each of six sessions, MN performed this task by imagining hand motion (see Methods) as soon as the target cue appeared. The task was performed immediately after filter building without intervening practice. Timing and directional tuning features of MI neurons during imagined actions were consistent with those observed in MI of intact non-human primates. Figure 4 shows that spike-rate modulation occurs soon after the 'go' cue and that modulation varied by target location, as would be predicted for MI if actual arm motions were performed[17]. Furthermore, 66 out of 73 discriminated units (90.4%) significantly changed their firing rate in relation to the appearance of the go cue (Kolmogorov–Smirnov test, $\alpha = 0.05$, rate calculated over a sliding 1-s window, overlapping every 0.05 s; 60-s data set for each condition, $n = 3$ sessions). These results indicate that, even years after spinal cord injury and in the absence of kinaesthetic feedback and limb movement, MI neurons can still be actively engaged and encode task-related information during the intention to move the limb ordinarily controlled by that MI region.

Quality of neural cursor control

Neural cursor position was significantly correlated with technician cursor position during the last block of the pursuit filter building task (x coordinate $r^2 = 0.56 \pm 0.18$ and y coordinate $r^2 = 0.45 \pm 0.15$, $n = 6$ sessions, Fig. 5). These correlations are similar or better than those seen in intact monkeys when linear filters were used to predict real-time hand position from MI neuronal ensembles[5,22]. The neural cursor could be directed towards targets with a form qualitatively similar to that seen for intact monkeys using closed-loop neural control[17,18,23]. As in intact monkeys, neural cursor motion had underlying instabilities and variable oscillatory components compared to hand motions of able-bodied individuals. Continuous neural cursor motion with the linear filter made cursor fixation at a single location difficult to achieve.

Data from the centre-out task were used to evaluate the speed and accuracy of cursor control, which are essential design parameters for any future practical NMP. As shown in Fig. 6, the participant correctly acquired 73–95% of targets (control 6.5%; $n = 80$, paired t-test, $P < 0.0001$, see Methods) when measured in a series of six sessions (see also Supplementary Fig. 2). Performance errors reflected both instabilities in cursor direction control and the ability to hold at the target location. Mean time to target was 2.51 ± 0.16 s ($\pm$s.e.m.) for successfully acquired targets. Although the best 13% of MN's trials were within the range consistently achieved by able-bodied controls using a computer mouse ($n = 3$, mean 1.06 ± 0.08 s ($\pm$s.e.m.)), the distribution of times for MN using neural control is skewed to longer acquisition times (Fig. 6b). Effective use of the neural cursor in more complex spatial control tasks was evident when MN directed the cursor to randomly placed targets while attempting to avoid obstacles in the cursor's path (see Supplementary Video 5).

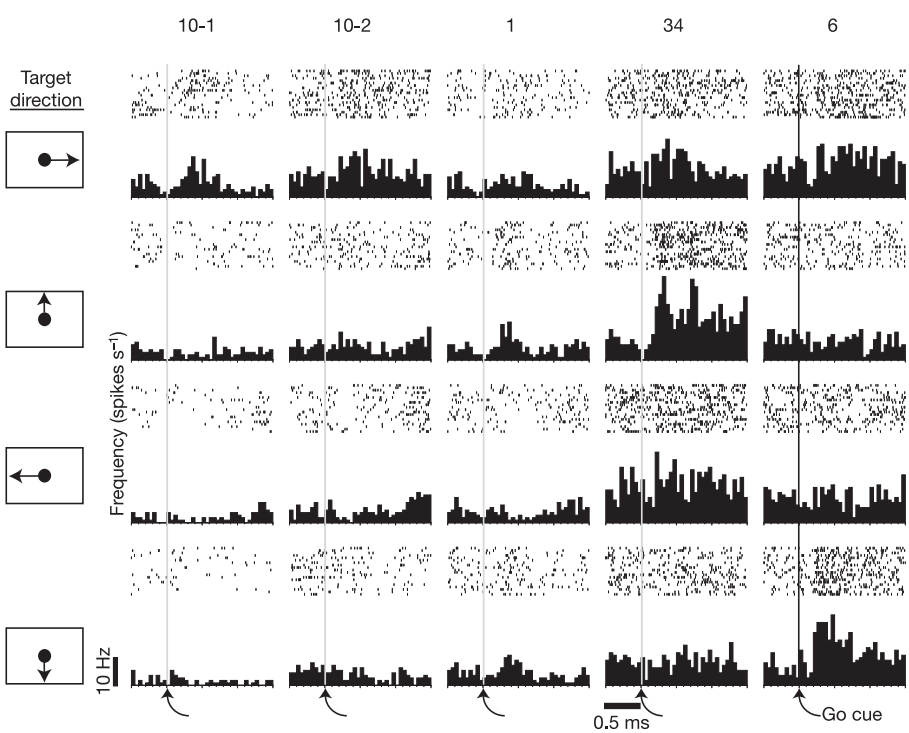


Figure 4 | Directional tuning during centre-out task. Peristimulus time histograms show spike rates for five neurons recorded simultaneously during the performance of a four-direction centre-out task (day 90) in which MN used the neural cursor to acquire a target presented at the right, top, left, or bottom of the screen. Twenty trials are displayed for each target location. Increases in activity after the go cue demonstrate movement-intention-related modulation. Each column shows the firing of one unit in the four directions, aligned on the cue to move. Note, for example, the time-locked increase in firing of unit 6 when MN was cued to move the cursor downward (lower right corner) and the lack of change in firing rate for upward instruction. Changes in firing across the five neurons reveal directional tuning.

© 2006 Nature Publishing Group

Further information concerning spatial and temporal accuracy was obtained from a 'grid task' (see Supplementary Information). As shown in Supplementary Video 8, participant 2 also performed the centre-out task, highlighting the ability for this second participant with spinal cord injury to modulate his motor cortical activity voluntarily for external device control. This participant's neural control, however, was generally less accurate and consistent than MN's; we are investigating the extent to which technical or other factors may underlie this performance. This and other participants' data will be reported in subsequent manuscripts.

Although he is tetraplegic, MN retains shoulder, neck and head mobility, and some recorded MI cells fired during shoulder movement (see Supplementary Fig. 1). NMPs will nearly always operate in the context of some existing movement capabilities (given the considerable variability of remaining sensory and motor functions in people with CNS injury); thus, during filter building and use MN was not asked to remain completely still, and he sometimes moved his head or neck. It is thus possible that retained movements influenced cursor action, much as such movements are known to influence activity correlated with ordinary hand actions[24]. These still-intact movements, however, did not appear to be essential for MN's cursor control, as can be appreciated particularly when he played 'Neural Pong' (Supplementary Video 4). In this video, movement of the head or the shoulders sometimes accompanied cursor control, but at other times MN moved the cursor purposefully while his head remained stationary, and during other epochs his head moved but the cursor motion appeared to be unrelated to head movement. We compared neural cursor position in Supplementary Video 4 to head position (both assessed by coordinates on a video playback monitor) and found no consistent relationship ($r^2 = 0.063$). This finding is inconsistent with a unique causal relation between head and cursor control. In addition, it was our repeated observation that, at least some of the time, MN performed neural control (and open-loop) tasks without moving his shoulder. Thus, based upon our instructions and MN's self-reporting that he was actively imagining arm action, we conclude that cursor motion is under the guidance of imagined/intended actions.

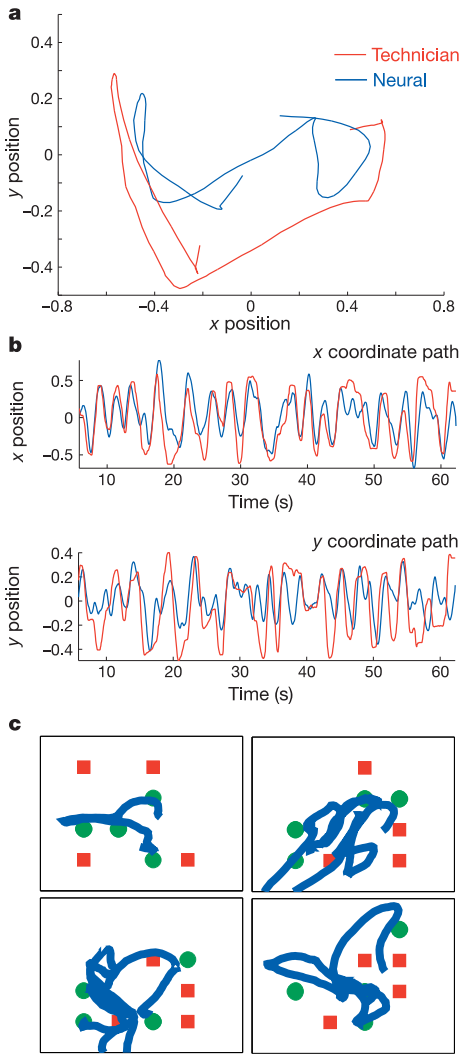


Figure 5 | Reconstruction of neural cursor position during pursuit tracking. **a**, Technician (red line) and neural cursor (blue line) paths during a 5-s epoch during which MN was asked to track the technician cursor with his neural cursor in real time. MN was able to track the general direction of the technician cursor with the neural cursor, changing directions quickly, while having some difficulty in overlaying the cursors precisely. Trial day 90. **b**, *x*, *y* position control over time during one tracking trial (last 1-min epoch of filter building). The top panel displays the *x* coordinate position of the technician cursor (red) and the neural cursor (blue); *y* positions for the same movement are shown in the bottom panel. **c**, Neural cursor position during a target acquisition/obstacle avoidance task. The four panels represent the four epochs shown in Supplementary Video 5. Green circles indicate targets; red squares indicate obstacles. The thick blue line indicates the path taken by the neural cursor and illustrates the ability to avoid most obstacles and acquire most targets within a randomly arranged field. Data are from trial day 90.

Direct control of prosthetic devices

Continuous computer cursor control could be used to provide many valuable new outputs for a person with paralysis to carry out

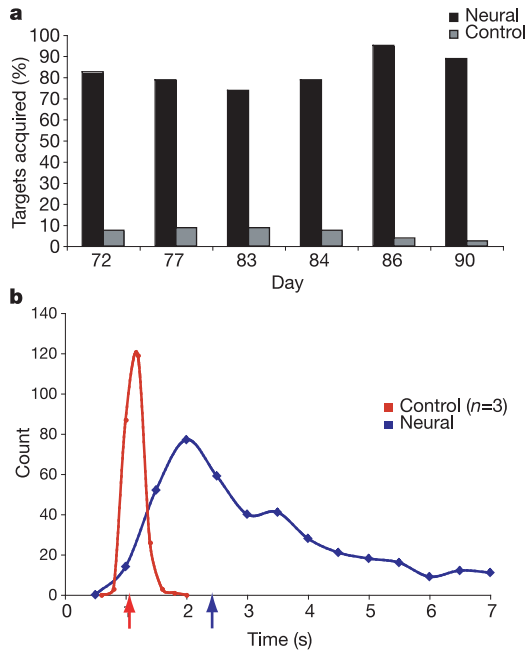


Figure 6 | Centre-out task performance. **a**, Target acquisition accuracy during the centre-out task. For each of six sessions, MN acquired between 73–95% of the radially placed targets. Control targets were not present on the monitor during task performance, but were marked as acquired if, during post-hoc analysis of the cursor movement, the cursor had traversed the location of one of the other three pseudo-randomly selected targets before the correct target (see Supplementary Video 1). Data from days 72, 77, 83, 84, 86, 90 are shown. **b**, Time-to-target performance during centre-out task for MN (blue) and three able-bodied controls (red). Only successful target acquisitions in <7 s are shown for MN. Arrows on the abscissa represent median times to target for each distribution. Controls' performances ($n = 3$ controls, 80 trials each) are collapsed into 0.2-s bins. MN's performance (398 trials) is collapsed into 0.5-s bins for visual clarity.

© 2006 Nature Publishing Group

activities of daily living. Such a control signal could be used not only to direct computer software, but also to operate external physical assistive devices. MN tested his ability to perform potentially useful actions in a series of demonstrations.

MN used a simplified computer interface to open simulated e-mail and to draw an approximately circular figure using a paint program (Supplementary Video 2). Using the neural cursor coupled to a simple hardware interface, he adjusted the volume, channel and power to his television. He was also able to play video games such as Neural Pong (see Supplementary Videos 3–5). Control of two robotic devices was achieved, allowing MN to manipulate directly the environment. In one example, neural output was coupled to a prosthetic hand (Liberating Technologies, Inc.; see Methods) and MN was able to open and close the hand under volitional control (see Supplementary Video 6). Although only a one-dimensional form of proportional control, he achieved this action after a few trials while looking at the hand and without requiring feedback from the cursor display. Lastly, MN used a simple multi-jointed robotic limb to grasp an object and transport it from one location to another (see Supplementary Video 7 and Supplementary Information). These demonstrations illustrate transfer of control directly to a physical device without depending on continued viewing of computer cursor feedback, and suggest that manipulation of the environment to enable activities such as self-paced eating, as well as other goal-directed actions, could be achieved in tetraplegic humans.

Notably, each of these tasks was achieved rapidly and could be performed while the participant was conversing. Thus, the MI-based NMP may have the property of allowing external device control with little more disruption than encountered in able-bodied humans when they are using their arms or hands and simultaneously carrying out other motor or cognitive functions.

Discussion

The research shown here from the first participant in an ongoing pilot clinical trial provides initial evidence that a human unable to move or sense his limbs can operate a NMP using MI neuronal ensemble spiking activity as a control source. In addition, we demonstrate that neural spiking remains in the MI arm area and can be modulated by intention years after spinal cord injury. These findings provide a number of novel insights concerning cortical function and the impact of spinal cord injury in humans (see Supplementary Information).

Although MN used direct neural control to perform reasonably challenging tasks, the level of the control is considerably less than that of an able-bodied person using a manually controlled computer mouse. A number of factors might affect control, including: (1) the small set of randomly selected neurons recorded by this single array, compared to the very large number usually engaged; (2) the influence of spinal cord injury mechanisms or duration since injury; (3) the cortical layers recorded; (4) our approach to filter building; (5) the nature of the linear filter (as compared, for example, to Kalman filters[22], adaptive algorithms or support vector machine approaches[25], not yet tested in our work); (6) attention and motivational state during filter building or control; and (7) the user interface. Changes in the recorded population across days may also contribute to both variability and instability of control. Shifting ensembles may result from small motions of the array or through other poorly understood mechanisms. Despite these variables, it is important to note that useful filters could be created daily from that neural population, and that advances in knowledge and technology are likely to improve recording and decoding. For example, cursor control might be enhanced further by adaptive algorithms or more selective choice of neural signals. Combinations of spiking activity and simultaneously recorded LFPs would provide additional signals that might improve performance.

Efforts to optimize performance may help to identify the design criteria for a clinically useful interface for paralysed humans. For example, electroencephalogram (EEG)-based control systems have been improved by limiting trial time, re-centring the cursor location after each trial, and stopping cursor motion as soon as targets were hit[26]. By contrast, our participant had to maintain cursor control at all times without these interface enhancements. However, we found that trial success was greater when target dwell time was decreased (grid task). Incorporation of additional feedback (for example, visual, auditory, somatosensory) may also be useful in enhancing performance[27].

Cursor and external device control may also be improved through learning[28]. MI of intact animals shows learning-related plasticity[29–32], and such plasticity undoubtedly contributes to the acquisition of direct cortical neuronal control of an NMP. Thus, we expect that learning will further improve control, as has been reported in monkeys[18]. The circle drawing task (Supplementary Video 2), for example, provides some evidence of improved performance with brief practice, but a better understanding of learning will require additional, more systematic investigation.

Other human BCIs are under development[33]. Foremost for comparison here is one by Kennedy and collaborators, who have reported results with their neurotrophic (cone) electrode for three humans: one with amyotrophic lateral sclerosis[34], one with brainstem stroke[35,36], and one with mitochondrial myopathy[36] (see Supplementary Information). In contrast to previous studies, here we present evidence that cortical ensemble patterns can be decoded and used for real-time, two-dimensional control of a computer cursor and other external devices by a person with spinal cord injury. Scalp-based EEG-driven BCIs also have been tested with some success in people with amyotrophic lateral sclerosis, spinal cord injury and other forms of paralysis[26,27,37–39]. Such indirect systems use a substitute for motor ensembles to achieve control. Although two-dimensional control has been achieved in paraplegic patients using independent modulation of scalp-recorded signals[26], this system requires significant learning, concentration to the exclusion of other actions, and daily scalp sensor application, but notably, not surgical placement of the sensor. It also has limited scalability to multiple functions because two-dimensional tasks appear to engage all controllable signals. In contrast, the direct NMP can be used during natural activities such as speech, and requires minimal learning beyond filter creation. It is also plausible that NMPs could be scaled so that parallel commands could be derived simultaneously from multiple sensors each in separate cortical regions. If achieved, relatively independent outputs potentially could emanate bilaterally from arm and leg areas to allow for the reanimation of paralysed limbs via functional electrical stimulation devices.

The relative merits of EEG, transcranial[40], electrocorticographic (ECoG)[41–43] and intracortically based BCIs (NMPs) continue to be explored and expanded[32,33]. Myriad issues including efficacy, safety (particularly with regards to surgery), reliability, longevity, required training and support, cosmesis, cost and availability will undoubtedly affect BCI success; individuals with impairments in communication and/or mobility may prioritize these factors differently when choosing a BCI, and it is possible that combinations of techniques will provide most useful restoration of control. Current BCIs (including the NMP reported here) depend upon the insights and assistance of trained experts. The need for this assistance must be eliminated through system automation.

The goal of NMPs and other BCI research is to create safe, reliable and unobtrusive neural interfaces to devices that will restore the communication, mobility, and independence of paralysed humans. NMPs could provide significant advances over current technologies because they engage natural substrates for control (that is, the MI arm/hand area), do not encumber other actions, and do not require extensive learning. It will be important to collect results from additional participants in order to understand the generality of these statements. Improvements in the decoder and user interface

© 2006 Nature Publishing Group

are likely to provide control closer to that achieved by able-bodied humans. The potential to use implanted sensors to 're-route' neural signals may also prove valuable in stroke and other neurological disorders where pathways are disconnected. The present NMP system incorporates a transcutaneous connection that tethers a participant to a bulky cart and requires operation by a trained technician. A wireless, implantable and miniaturized system combined with automation will be required for practical use. Emerging and available technologies appear to be sufficient to overcome these obstacles, although the challenges of creating a fully implantable system may be formidable. The current data provide initial proof of concept that spiking activity from neuronal ensembles can provide a control signal after spinal cord injury sufficient to perform at least basic operations for a human with tetraplegia, justifying further engineering efforts towards a human NMP.

METHODS

See Supplementary Information for additional Methods.

Approval for these studies was granted by the US Food and Drug Administration (Investigational Device Exemption) and the Rhode Island Hospital, New England, Rehabilitation Institute of Chicago, University of Chicago, and Spaulding Rehabilitation Hospital Institutional Review Boards. The participant described in this report has provided permission for photographs, videos and portions of his protected health information to be published for scientific and educational purposes. After completion of informed consent and medical and surgical screening procedures, the 4 × 4-mm array of electrodes was implanted into the motor cortex using a pneumatic insertion technique[8,44]. Details of the human surgical procedure are in preparation for publication.

BrainGate system. The sensor is a 10 × 10 array of silicon microelectrodes that protrude 1 mm from a 4 × 4-mm platform (Fig. 1b). At manufacture, electrodes had an impedance of 322 ± 138 kOhm (mean ± s.d.). The array was implanted onto the surface of the MI arm/hand region; electrodes penetrate into the cortex to attempt to record neurons in intermediate layers. Recorded electrical signals pass externally through a Ti percutaneous connector, which is secured to the skull. Cabling attached to the connector during recording sessions routes signals to external amplifiers and then to a series of computers in a cart that process the signals and convert them into an output (the neural cursor) that can be viewed by the participant on a computer monitor (Fig. 1d). Currently, the system must be set up and managed by an experienced technician.

Recording sessions. Research sessions were scheduled at least once per week at the participant's residence. Sessions could be cancelled (for example, due to participant request or schedule conflict) or ended early at the participant's request. Sessions would commence with neural recording and spike discrimination, followed by filter building and structured clinical end-point (cursor control) trials. Performance of video games and external device control demonstrations followed. The electrodes and neural signals selected immediately before filter building remained constant for any given session's cursor control trials; additional neural signals (detected either with alternative spike discrimination methods, or as may have become evident during the cursor control trials) could be used during subsequent video game or external device use.

Spike discrimination. Units were manually discriminated by a technician using visual features to place time-amplitude windows on waveforms displayed within 1.6-ms windows triggered when the signal crossed a manually adjusted threshold (Fig. 2) (Cyberkinetics Central Software) while the participant was at rest. We applied no objective criteria to determine whether any particular unit (or admixture of units) recorded on one day was the same as that recorded on a subsequent day. Smaller amplitude signals were likely to have consisted of admixtures of spikes from >1 neuron. These signals were nevertheless used for sessions.

Filter building. For each session, single and multiunit data were used to create a filter to generate a two-dimensional output signal. During filter building, MN was asked to imagine moving his hand as if he were controlling a computer mouse. Specifically, he was asked to imagine tracking a cursor on the computer screen; this technician cursor was moved (by the technician) through a succession of randomly positioned targets. During the first 4 min of imagined tracking, only the technician cursor and targets were visible on the screen. An initial filter was built with the neural activity collected during this epoch, then allowing a neural cursor to be placed on the screen. MN continued to track the technician cursor with his neural cursor over a series of four 1-min blocks. The filter was updated at the end of each block. To create the final filter for the research session, neural data were used from only the four neural cursor blocks.

Linear filters were constructed from a response matrix containing the firing rate over a 1-s history for each neuron (twenty 50-ms bins), and regressing this matrix onto technician-cursor position using a pseudoinverse technique. The least-squares formulation comprises a closed-form solution:

$$\mathbf{u} = \mathbf{R}\cdot\mathbf{f} = \mathbf{R}(\mathbf{R}^{\mathrm{T}}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{T}}\mathbf{k}$$

where **R** is the neural response matrix, **f** is the linear filter, **k** is the kinematic value (x,y position), **T** indicates the transpose of the matrix, and **u** is the reconstruction[5]. Software for displays was custom generated for this purpose and run on standard PC computers (Orbit Micro, dual Intel Pentium 4 Xeon processors, 2 GHz).

External devices. The robot arm was obtained from Lynxmotion, Lynx 5 Series. The electric prosthetic hand was generously provided by Liberating Technologies.

Received 22 March; accepted 6 June 2006.

1. Jackson, A. B., Dijkers, M., Devivo, M. J. & Poczatek, R. B. A demographic profile of new traumatic spinal cord injuries: change and stability over 30 years. *Arch. Phys. Med. Rehabil.* **85,** 1740–1748 (2004).
2. Thoumie, P. *et al.* Clinical and functional evaluation of a gaze controlled system for the severely handicapped. *Spinal Cord* **36,** 104–109 (1998).
3. Humphrey, D. R., Schmidt, E. M. & Thompson, W. D. Predicting measures of motor performance from multiple cortical spike trains. *Science* **170,** 758–762 (1970).
4. Schwartz, A. B., Taylor, D. M. & Tillery, S. I. Extraction algorithms for cortical control of arm prosthetics. *Curr. Opin. Neurobiol.* **11,** 701–707 (2001).
5. Paninski, L., Fellows, M. R., Hatsopoulos, N. G. & Donoghue, J. P. Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *J. Neurophysiol.* **91,** 515–532 (2004).
6. Maynard, E. M., Nordhausen, C. T. & Normann, R. A. The Utah intracortical electrode array: a recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.* **102,** 228–239 (1997).
7. Guillory, K. S. & Normann, R. A. A 100-channel system for real time detection and storage of extracellular spike waveforms. *J. Neurosci. Methods* **91,** 21–29 (1999).
8. Suner, S., Fellows, M. R., Vargas-Irwin, C., Nakata, K. & Donoghue, J. P. Reliability of signals from chronically implanted, silicon-based electrode array in non-human primate primary motor cortex. *IEEE Trans. Neural Syst. Rehabil. Eng.* **13,** 524–541 (2005).
9. Maynard, F. M. Jr *et al.* International standards for neurological and functional classification of spinal cord injury. American Spinal Injury Association. *Spinal Cord* **35,** 266–274 (1997).
10. Yousry, T. A. *et al.* Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. *Brain* **120,** 141–157 (1997).
11. Lotze, M., Laubis-Herrmann, U., Topka, H., Erb, M. & Grodd, W. Reorganization in the primary motor cortex after spinal cord injury—A functional magnetic resonance (fMRI) study. *Restor. Neurol. Neurosci.* **14,** 183–187 (1999).
12. Shoham, S., Halgren, E., Maynard, E. M. & Normann, R. A. Motor-cortical activity in tetraplegics. *Nature* **413,** 793 (2001).
13. Sabbah, P. *et al.* Sensorimotor cortical activity in patients with complete spinal cord injury: a functional magnetic resonance imaging study. *J. Neurotrauma* **19,** 53–60 (2002).
14. Mikulis, D. J. *et al.* Adaptation in the motor cortex following cervical spinal cord injury. *Neurology* **58,** 794–801 (2002).
15. Donoghue, J. P., Sanes, J. N., Hatsopoulos, N. G. & Gaal, G. Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements. *J. Neurophysiol.* **79,** 159–173 (1998).
16. Sanes, J. N. & Donoghue, J. P. Plasticity and primary motor cortex. *Annu. Rev. Neurosci.* **23,** 393–415 (2000).
17. Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R. & Donoghue, J. P. Instant neural control of a movement signal. *Nature* **416,** 141–142 (2002).
18. Taylor, D. M., Tillery, S. I. & Schwartz, A. B. Direct cortical control of 3D neuroprosthetic devices. *Science* **296,** 1829–1832 (2002).
19. Evarts, E. V. in *Handbook of Physiology* (ed. Brooks, V.) 1083–1120 (Williams and Wilkins, Baltimore, 1981).
20. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2,** 1527–1537 (1982).
21. Humphrey, D. R. & Tanji, J. in *Motor Control: Concepts and Issues* (eds Humphrey, D. R. & Freund, H. J.) 413–443 (John Wiley, London, 1991).
22. Wu, W. *et al.* Modeling and decoding motor cortical activity using a switching Kalman filter. *IEEE Trans. Biomed. Eng.* **51,** 933–942 (2004).
23. Carmena, J. M. *et al.* Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* **1,** E42 (2003).
24. Baker, J. T., Donoghue, J. P. & Sanes, J. N. Gaze direction modulates finger movement activation patterns in human cerebral cortex. *J. Neurosci.* **19,** 10044–10052 (1999).
25. Olson, B. P., Si, J., Hu, J. & He, J. Closed-loop cortical control of direction using support vector machines. *IEEE Trans. Neural Syst. Rehabil. Eng.* **13,** 72–80 (2005).
26. Wolpaw, J. R. & McFarland, D. J. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl Acad. Sci. USA* **101,** 17849–17854 (2004).

© 2006 Nature Publishing Group

27. Birbaumer, N. *et al.* A spelling device for the paralysed. *Nature* **398,** 297–298 (1999).
28. Fetz, E. E. & Baker, M. A. Operantly conditioned patterns on precentral unit activity and correlated responses in adjacent cells and contralateral muscles. *J. Neurophysiol.* **36,** 179–204 (1973).
29. Fetz, E. E. & Finocchio, D. V. Operant conditioning of specific patterns of neural and muscular activity. *Science* **174,** 431–435 (1971).
30. Jacobs, K. M. & Donoghue, J. P. Reshaping the cortical motor map by unmasking latent intracortical connections. *Science* **251,** 944–947 (1991).
31. Gandolfo, F., Li, C., Benda, B. J., Schioppa, C. P. & Bizzi, E. Cortical correlates of learning in monkeys adapting to a new dynamical environment. *Proc. Natl Acad. Sci. USA* **97,** 2259–2263 (2000).
32. Schwartz, A. B. Cortical neural prosthetics. *Annu. Rev. Neurosci.* **27,** 487–507 (2004).
33. Hochberg, L. R. & Donoghue, J. P. Sensors for brain-computer interfaces. *IEEE Eng. Med. Biol.* (in the press).
34. Kennedy, P. R. & Bakay, R. A. Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport* **9,** 1707–1711 (1998).
35. Kennedy, P. R., Bakay, R. A., Moore, M. M., Adams, K. & Goldwaithe, J. Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.* **8,** 198–202 (2000).
36. Kennedy, P. R., Kirby, M. T., Moore, M. M., King, B. & Mallory, A. Computer control using human intracortical local field potentials. *IEEE Trans. Neural Syst. Rehabil. Eng.* **12,** 339–344 (2004).
37. Mason, S. G., Bohringer, R., Borisoff, J. F. & Birch, G. E. Real-time control of a video game with a direct brain–computer interface. *J. Clin. Neurophysiol.* **21,** 404–408 (2004).
38. Vaughan, T. M. *et al.* Brain-computer interface technology: a review of the Second International Meeting. *IEEE Trans. Neural Syst. Rehabil. Eng.* **11,** 94–109 (2003).
39. Muller-Putz, G. R., Scherer, R., Pfurtscheller, G. & Rupp, R. EEG-based neuroprosthesis control: a step towards clinical practice. *Neurosci. Lett.* **382,** 169–174 (2005).
40. Kennedy, P. *et al.* Using human extra-cortical local field potentials to control a switch. *J. Neural Eng.* **1,** 72–77 (2004).
41. Leuthardt, E. C., Schalk, G., Wolpaw, J. W., Ojemann, J. G. & Moran, D. W. A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.* **1,** 63–71 (2004).
42. Graimann, B., Huggins, J. E., Levine, S. P. & Pfurtscheller, G. Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis. *IEEE Trans. Biomed. Eng.* **51,** 954–962 (2004).
43. Lal, T. N., *et al.* in *Advances in Neural Information Processing Systems* (eds Saul, L. K., Weiss, Y. & Bottou, L.) (MIT Press, Cambridge, Massachusetts, 2005).
44. Rousche, P. J. & Normann, R. A. A method for pneumatically inserting an array of penetrating electrodes into cortical tissue. *Ann. Biomed. Eng.* **20,** 413–422 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank J. Joseph and D. Morris for assistance; L. Mermel for clinical planning advice; V. Zerris and M. Park for surgical assistance; G. Polykoff for clinical trial assistance; W. Truccolo for power spectral density analysis development; and the employees of Cyberkinetics for device engineering, manufacturing and clinical trial design and management. The authors also thank MN for his participation in this trial, and the nursing staff at his assisted care facility for their assistance. The authors are grateful to M. Serra and Sargent Rehabilitation Center, the study site, for administrative support. The photograph of MN (Fig. 1) is copyright 2005 Rick Friedman. This work was supported by Cyberkinetics Neurotechnology Systems, Inc.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to J.P.D. (john_donoghue@brown.edu).

© 2006 Nature Publishing Group

LETTERS

Energy input and response from prompt and early optical afterglow emission in γ-ray bursts

W. T. Vestrand[1], J. A. Wren[1], P. R. Wozniak[1], R. Aptekar[2], S. Golentskii[2], V. Pal'shin[2], T. Sakamoto[3], R. R. White[1], S. Evans[1], D. Casperson[1] & E. Fenimore[1]

The taxonomy of optical emission detected during the critical first few minutes after the onset of a γ-ray burst (GRB) defines two broad classes: prompt optical emission correlated with prompt γ-ray emission[1], and early optical afterglow emission uncorrelated with the γ-ray emission[2]. The standard theoretical interpretation attributes prompt emission to internal shocks in the ultra-relativistic outflow generated by the internal engine[3–5]; early afterglow emission is attributed to shocks generated by interaction with the surrounding medium[6–8]. Here we report on observations of a bright GRB that, for the first time, clearly show the temporal relationship and relative strength of the two optical components. The observations indicate that early afterglow emission can be understood as reverberation of the energy input measured by prompt emission. Measurements of the early afterglow reverberations therefore probe the structure of the environment around the burst, whereas the subsequent response to late-time impulsive energy releases reveals how earlier flaring episodes have altered the jet and environment parameters. Many GRBs are generated by the death of massive stars that were born and died before the Universe was ten per cent of its current age[9,10], so GRB afterglow reverberations provide clues about the environments around some of the first stars.

On 20 August 2005 a relatively faint 30-s pulse of γ-ray emission from GRB 050820a started at 06:34:53 Universal Time (UT), and was localized[11] in real time by the autonomous software of the Burst Alert Telescope (BAT) on the Swift satellite[12]. One of our autonomous RAPTOR (Rapid Telescopes for Optical Response) telescopes[13] began optical imaging of the GRB 050820a location 5.5 s (ref. 14) after distribution of the Swift alert. The images show emergence of faint optical emission that suddenly flares (see Fig. 1) and varies erratically before gradually fading over the course of the first hour. Comparison of the optical measurements with the γ-ray light curve measured by the KONUS-Wind experiment (Fig. 2) indicates that the rapid optical flares are simultaneous with the major outbursts of γ-ray emission. These data clearly confirm the idea that one component of the early optical light from GRBs is prompt emission that closely tracks the γ-ray emission[1]. The remaining optical emission persists even after the γ-rays disappear and can naturally be explained as an early afterglow component[2,15–18].

The optical light curve rise to maximum light for even the afterglow component has too much curvature (positive and negative) to be fitted by the self-similar power-law rise predicted by internal or external shock models. To dissect the light curve into primary components, we made the simple conjecture that the two optical components—prompt optical $F_p(t)$ and early optical afterglow $F_a(t)$ fluxes—have the temporal behaviours:

$$F_p(t) = C_p F_\gamma(t) \qquad (1)$$

$$F_a(t) = C_a \left(\frac{t - t_o}{t_o}\right)^{-s} \exp\left(\frac{-\tau}{t - t_o}\right) \qquad (2)$$

Here $F_\gamma(t)$ is the prompt γ-ray flux; t_o is the time for onset of energy release; τ is the timescale for rise of the afterglow; s is the power-law

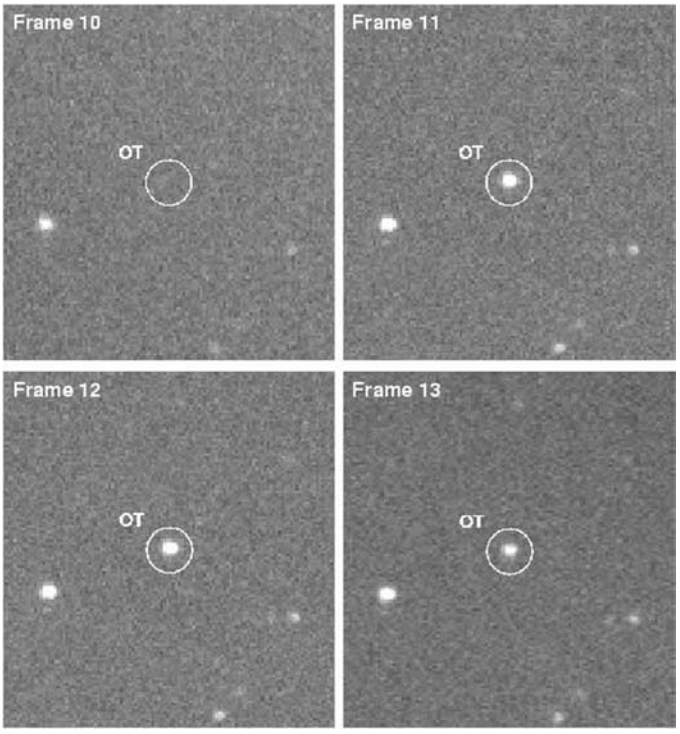


Figure 1 | The onset of prompt optical emission from GRB 050820a. The images show the GRB location in four consecutive frames collected by the 0.4-m fully autonomous rapid-response telescope, RAPTOR-S, owned by the Los Alamos National Laboratory and located at an altitude of 2,500 m in the Jemez Mountains of New Mexico. The RAPTOR-S telescope uses an unfiltered 1,000 × 1,000 pixel back-illuminated charge-coupled device (CCD) camera and typically achieves a 5-sigma limiting magnitude of R ≈ 19 mag for 30-s exposures. The displayed images span the time interval 06:38:07.2 to 06:40:50.7 UT on 20 August 2005 and the circle denotes the location of the GRB optical counterpart (right ascension 22 h 29 min 38.1 s, declination +19° 33′ 37.1″ (J2000))[26] that subsequent spectral measurements[27,28] showed was at a redshift of $z = 2.612 \pm 0.002$. The top two images (frames 10 and 11) show flaring of the optical transient (OT) by more than two magnitudes in less than 30 s. The bottom two images (frames 12 and 13) are the next two 30-s exposures that show an abrupt drop of the optical emission in frame 13 by ~0.5 magnitudes.

[1]Los Alamos National Laboratory, Space Science and Applications Group, ISR-1, MS-D466, Los Alamos, New Mexico 87545, USA. [2]Ioffe Physico-Technical Institute, St Petersburg, 194021, Russia. [3]NASA Goddard Space Flight Center, Code 661, Greenbelt, Maryland 20771, USA.

© 2006 Nature Publishing Group

decay index; and C_p, C_a are the relative amplitudes of the prompt and early afterglow components, respectively. After re-binning the γ-ray measurements to the same observing intervals as the optical observations (see Table 1), we find that the simple temporal behaviours given by equations (1) and (2) jointly describe the optical light curve measured for GRB 050820a rather well (see Fig. 3).

The prompt optical component given by equation (1) reproduces the fast variations observed in the RAPTOR light curve when the KONUS γ-ray measurements are scaled by the flux ratio $C_p = F_{opt}/F_\gamma \approx 7 \times 10^{-6}$—a ratio is comparable to the value of $F_{opt}/F_\gamma = 1.2 \times 10^{-5}$ found for GRB 041219a[1]. For both events there

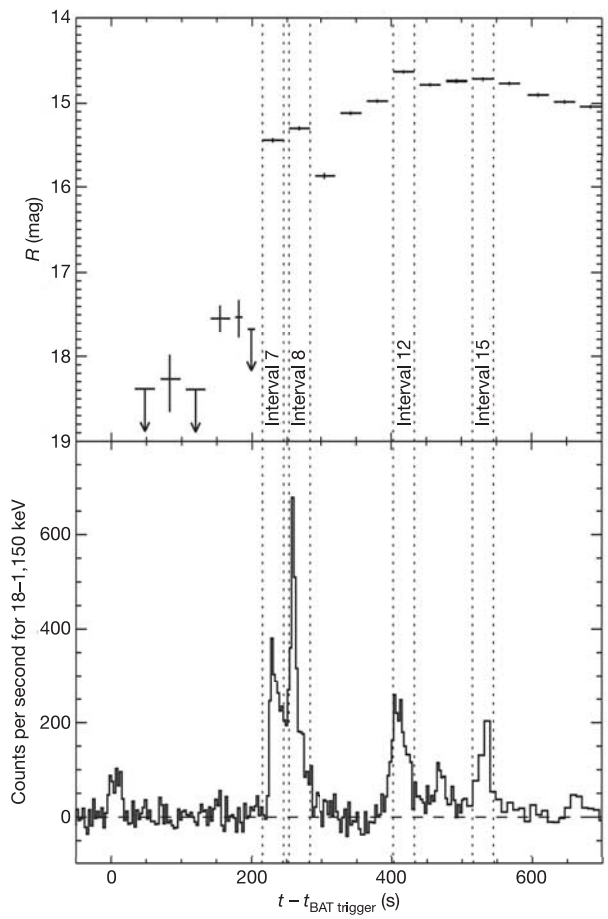


Figure 2 | A comparison of the early optical light curve and the γ-ray light curve measured for GRB 050820a. The lower panel shows the γ-ray count rate in the light curve measured by KONUS γ-ray detectors on board the WIND satellite. The upper panel shows the optical light curve measured by the RAPTOR-S telescope. The horizontal lines in the upper panel denote the durations of the observation intervals and the vertical lines denote the 1-sigma error bars associated with the optical magnitude measurements made during those intervals. After that initial pulse, the GRB was quiescent at γ-ray energies for more than three minutes until a major outburst began at 06:38:40 UT. The observations by Swift were truncated by satellite passage into a high background region, but the KONUS γ-ray detector aboard the Wind satellite was able to measure the entire complex γ-ray light curve[29], which lasted ~750 s. Integrated over the entire event, the total (18–1,000 keV) γ-ray fluence for GRB 050820a was 5.3×10^{-5} erg cm^{-2}, which for the observed redshift[27,28] ($z = 2.6$) and standard cosmological parameters ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1}) corresponds to an isotropic energy of 8.1×10^{53} erg. The abrupt brightening of the optical emission shown in Fig. 1 occurs simultaneously with the major pulses of γ-ray emission shown in the interval between 226 and 290 s after the burst trigger. The onset of early optical afterglow emission is visible in the interval between 300 and 400 s after the burst trigger. $t_{BAT\ trigger}$, time that BAT was triggered.

Table 1 | Joint RAPTOR/KONUS observations of GRB 050820a

Interval	t_{start} (s)	t_{end} (s)	Δt_{exp} (s)	R (mag)	F (18–1,000 keV) (10^{-7} erg cm^{-2})
0	−4.30	19.30	23.6	–	25.9 ± 1.3
1	33.72	61.87	30	>18.38	<3.3
2	70.08	98.27	30	18.270 ± 0.290	<4.5
3	105.97	133.56	30	>18.39	<6.0
4	141.16	168.76	20	17.540 ± 0.150	<5.6
5	176.35	186.35	10	17.530 ± 0.200	<2.2
6	194.05	204.05	10	>17.67	<3.7
7	214.85	244.85	30	15.437 ± 0.025	101.9 ± 1.4
8	252.44	282.44	30	15.297 ± 0.024	160.2 ± 1.5
9	290.04	320.04	30	15.870 ± 0.033	16.6 ± 1.2
10	327.63	357.63	30	15.115 ± 0.022	<5.1
11	365.23	395.23	30	14.972 ± 0.020	7.8 ± 1.1
12	420.83	432.83	30	14.633 ± 0.019	87.7 ± 1.2
13	440.42	470.42	30	14.779 ± 0.019	27.7 ± 1.2
14	478.11	508.11	30	14.740 ± 0.020	21.5 ± 1.1
15	515.71	545.71	30	14.718 ± 0.019	51.5 ± 1.1
16	553.30	583.30	30	14.764 ± 0.019	11.5 ± 1.1
17	594.20	624.20	30	14.901 ± 0.020	4.7 ± 3.3
18	631.79	661.79	30	14.983 ± 0.020	<5.4
19	669.39	699.39	30	15.038 ± 0.021	11.9 ± 1.1
20	706.98	736.98	30	15.060 ± 0.022	11.5 ± 1.3

The R-band magnitudes were derived by transforming the unfiltered measurements to an R-band equivalent using the USNO-B1.0 magnitudes for comparison stars in the RAPTOR images.

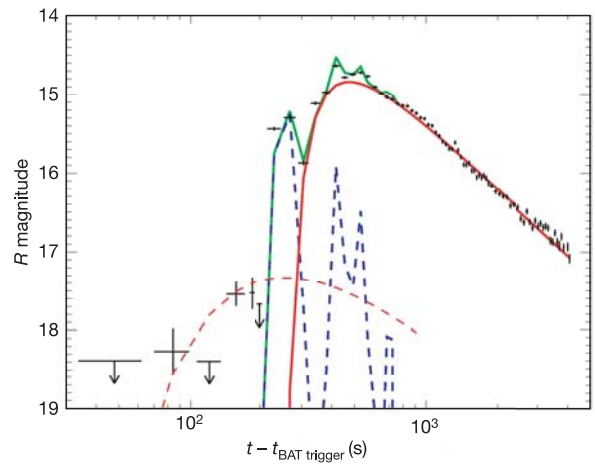


Figure 3 | The decomposition the optical light curve measured for GRB 050820a into primary optical components. The R-band magnitudes measured by RAPTOR-S are indicated by the black crosses, with observing intervals denoted by horizontal lines and 1-sigma error bars represented by the vertical lines. The blue dashed trace shows the prompt optical component obtained by scaling the KONUS γ-ray flux measurements, after re-binning to the same time intervals as the RAPTOR measurements, by the factor $F_{opt}/F_\gamma = 7.4 \times 10^{-6}$ and converting to the R-band equivalent magnitude. The solid red line shows the model early afterglow component of the form given by equation (2) with a flux-rise timescale of $\tau \approx 280$ s, the late-time power-law flux decay with index $s \approx 1.1$, and a reference time t_o equal to the start of the dominant γ-ray outburst (226 s after the BAT trigger time). The green trace shows the sum of these prompt and early afterglow model components. The dashed red line shows an afterglow with the same flux-rise timescale as the dominant afterglow component ($\tau \approx 280$ s) but with a reference time appropriate for the precursor pulse ($t_o = 5$ s) and an amplitude given by the ratio of the γ-ray fluence for the precursor pulse to that of the main (18–1,150 keV) pulse of 0.1. We note that even when the prompt γ-ray-emitting intervals are dropped, the rise of the afterglow emission cannot be fitted with a power-law dependence. By fitting a power law with a reference time of $t_o = 0.0$ to the fast rise in the interval, measured with unprecedented signal-to-noise between 300 and 400 s after the trigger, we predict fluxes that are inconsistent with the detections of optical emission by RAPTOR and the UVOT in the interval before 200 s after the trigger.

© 2006 Nature Publishing Group

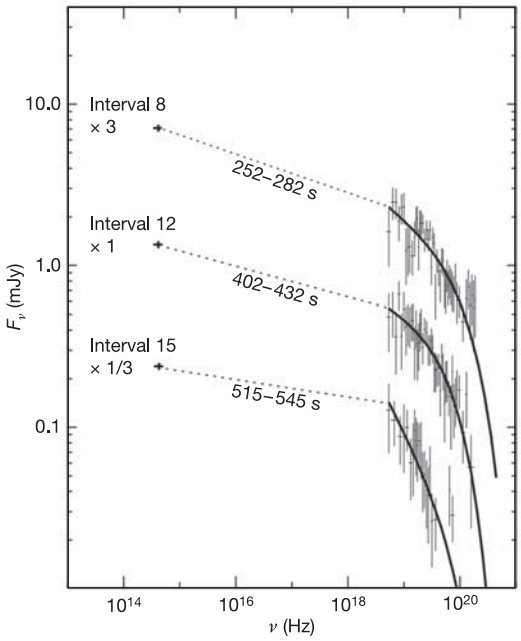


Figure 4 | Broadband spectra of prompt emission from GRB 050820a. These spectra are constructed from simultaneous optical measurements from the RAPTOR-S telescope and γ-ray measurements from the KONUS experiment on board the WIND satellite. The crosses at the low frequencies denote the optical flux density after correction for the contribution from early optical afterglow emission. The solid line denotes the best-fitting model for the KONUS measurements and the high-frequency crosses represent the individual channel measurements with the 1-sigma errors. The time intervals are measured in seconds from the Swift trigger time. During intervals 8 and 12, the broadband spectra of prompt emission for GRB 050820a show optical flux density levels that are roughly consistent with an extrapolation of the high-energy spectral shape. But in interval 15, extrapolation of the γ-ray spectrum underpredicts the optical flux. This suggests that the prompt optical and γ-ray emission might be generated by separate, but closely linked, radiation processes.

are intervals in which the broadband prompt spectra (Fig. 4) between optical and γ-ray bands must on average be flatter than the $F_\nu \propto \nu^{-1/2}$ expected for the standard synchrotron picture with fast cooling electrons and a cooling critical frequency near the optical band[7]. A possible explanation for closely linked, but separate, spectral components is that they are both associated with the internal jet shock, but that optical emission is generated by the reverse shock and γ-rays by the forward shock[7]. This scenario would place bounds on the jet bulk velocity, the overtaking shell thickness, and the distance at which the emission originates. An alternate internal shock explanation is that the optical light is synchrotron emission and the γ-rays are inverse-Compton-scattered synchrotron photons[4]. The closer tracking of the optical flux with the highest-energy γ-ray band also seen in GRB 041219a is naturally explained in this scenario. Detailed modelling work will be needed to test the standard synchrotron picture and, if necessary, discriminate between possible alternative models, but the close temporal tracking of the prompt components will place important constraints on the jet properties, which are difficult to obtain in other ways.

The dramatic optical flux increase immediately after the dominant γ-ray pulse suggests that the afterglow is associated with the impulsive energy release signalled by the prompt emission. After the rapid initial rise, the rate of flux increase slows and transitions into a shallow decline that gradually steepens to a power-law flux decay with an index of $s = 1.1$ after $\sim 10^3$ seconds. Similar late-time behaviour has been observed for several other events[18,19] and is consistent with that expected for an external forward shock in the internal/external shock paradigm.

Although neither instrument was observing the GRB location during the precursor pulse that triggered Swift/BAT, both RAPTOR and the Ultra-Violet and Optical Telescope (UVOT) on Swift did detect[20] faint optical emission during the interval after the precursor and before the outburst of intense γ-ray emission. During that same interval neither Swift nor KONUS detected γ-ray emission. The best explanation for this faint optical emission is that it is afterglow emission from the precursor γ-ray pulse—an explanation supported by the Swift detection of a fading soft X-ray afterglow during the same interval[21]. Further, the RAPTOR measurements of this precursor afterglow show a light curve shape that is consistent with a straightforward scaling of the main afterglow (Fig. 3).

Until now, observational determination of the light curve shape has been hampered by not knowing where to place the onset reference time t_o (refs 22, 23). In other words, when does the afterglow begin? Often it is placed at the burst trigger time, but it can be placed logically anywhere within, or even before, the γ-ray-emitting interval. The exact choice can substantially modify the derived shape of the early light curve[22,23]. Our observations show: (1) that the onset of the dominant afterglow component should be referenced to the time of the onset of the dominant γ-ray pulse; and (2) that the assumption of a single onset time for the afterglow, t_o, is too simple. The structure of the afterglow is better understood as a reverberation that can be related to the stimulus measured by the prompt emission convolved with a transfer function representing the response of the system.

There is growing evidence that the GRB engine can impulsively release energy well after the initial explosion[24,25]. Measurements of the broad-band spectra of the prompt emission place important constraints on the evolution of the jet itself. Structure associated with the afterglow from those secondary energy releases can also emerge as the primary afterglow component fades. The timing and strength of those secondary reverberations will probe the evolution of the interaction and how the GRB environment is modified. Long-duration GRBs are known to signal the deaths of massive stars and occur at very high redshift[9,10], so measurement of GRB reverberation and its temporal variations can be used to map out the nurseries of the earliest stars.

Received 1 February; accepted 15 May 2006.

1. Vestrand, W. T. *et al.* A link between prompt optical and prompt γ-ray emission in γ-ray bursts. *Nature* **435**, 178–180 (2005).
2. Akerlof, C. *et al.* Observations of contemporaneous optical radiation from a γ-ray burst. *Nature* **398**, 400–402 (1999).
3. Fenimore, E. E., Madras, C. D. & Nayakshin, S. Expanding relativistic shells and gamma-ray burst temporal structure. *Astrophys. J.* **473**, 998–1012 (1996).
4. Meszaros, P. & Rees, M. GRB 990123: reverse and internal shock flashes and late afterglow behaviour. *Mon. Not. R. Astron. Soc.* **306**, L39–L43 (1999).
5. Katz, J. Low-frequency spectra of gamma-ray bursts. *Astrophys. J.* **432**, L107–L109 (1994).
6. Meszaros, P. & Rees, M. Optical and long-wavelength afterglow from gamma-ray bursts. *Astrophys. J.* **476**, 232–237 (1997).
7. Sari, R. & Piran, T. Predictions for the very early afterglow and the optical flash. *Astrophys. J.* **520**, 641–649 (1999).
8. Panaitescu, A. & Kumar, P. Analysis of two scenarios for the early optical emission of the gamma-ray burst afterglow 990123 and 021211. *Mon. Not. R. Astron. Soc.* **353**, 511–522 (2004).
9. Lamb, D. Q. & Reichart, D. E. Gamma-ray bursts as a probe of the very high redshift universe. *Astrophys. J.* **536**, 1–18 (2000).
10. Ciardi, B. & Loeb, A. Expected number and flux distribution of gamma-ray burst afterglows with high redshifts. *Astrophys. J.* **540**, 687–696 (2005).
11. Page, M. *et al.* GRB 050820: Swift detection of a GRB. *GRB Circ. Netw.* **3830** (2005).
12. Gehrels, N. *et al.* The Swift gamma-ray burst mission. *Astrophys. J.* **611**, 1005–1020 (2004).
13. Vestrand, W. T. *et al.* The RAPTOR experiment: a system for monitoring the optical sky in real time. *Proc. SPIE* **4845**, 126–136 (2002).
14. Wren, J. *et al.* GRB050820: Early RAPTOR detections. *GRB Circ. Netw.* **3836** (2005).
15. Van Paradijs, J., Kouveliotou, C. & Wijers, R. A. M. J. Gamma ray burst afterglows. *Annu. Rev. Astron. Astrophys.* **38**, 379–425 (2000).
16. Li, W. *et al.* The early light curve of the optical afterglow of GRB 021211. *Astrophys. J.* **586**, L9–L12 (2003).

© 2006 Nature Publishing Group

17. Fox, D. W. *et al.* Early optical emission from the γ-ray burst of 4 October 2002. *Nature* **422,** 284–286 (2003).
18. Wozniak, P. R. *et al.* RAPTOR observations of the early optical afterglow from GRB 050319. *Astrophys. J.* **627,** L13–L16 (2005).
19. Rykoff, E. S. *et al.* The early optical afterglow of GRB 030418 and progenitor mass loss. *Astrophys. J.* **601,** 1013–1018 (2004).
20. Chester, M. *et al.* Swift/UVOT photometry of GRB 050820. *GRB Circ. Netw.* **3838** (2005).
21. Page, K. L. *et al.* GRB 050820: refined XRT analysis. *GRB Circ. Netw.* **3837** (2005).
22. Lazzati, D. & Begelman, M. Thick fireballs and the steep decay in the early X-ray afterglow of gamma-ray bursts. *Astrophys. J.* **641,** 972–977 (2006).
23. Quimby, R. M. *et al.* Early-time observations of the GRB 050319 optical transient. *Astrophys. J.* **640,** 402–406 (2006).
24. Falcone, A. D. *et al.* The giant x-ray flare of GRB 050502b: evidence for late-time internal engine activity. *Astrophys. J.* **641,** 1010–1017 (2006).
25. Wozniak, P. R. *et al.* RAPTOR observations of delayed explosive activity in the high-redshift gamma-ray burst GRB 060206. *Astrophys. J.* **642,** L99–L102 (2006).
26. Fox, D. B. & Cenko, S. B. GRB 050820: Optical afterglow from P60. *GRB Circ. Netw.* **3829** (2005).
27. Prochaska, J. X. *et al.* GRB 050820: High resolution spectroscopy from Keck. *GRB Circ. Netw.* **3833** (2005).
28. Ledoux, C. *et al.* VLT/UVES spectroscopy of GRB050820. *GRB Circ. Netw.* **3869** (2005).
29. Golenetskii, S. *et al.* GRB050820a—a very long GRB like GRB041219a? Konus-Wind Observation. *GRB Circ. Netw.* **3852** (2005).

Acknowledgements The RAPTOR project is supported by the Laboratory Directed Research and Development program at Los Alamos National Laboratory. The Konus-Wind experiment is supported by the Russian Space Agency and the Russian Foundation for Basic Research.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.T.V. (vestrand@lanl.gov).

© 2006 Nature Publishing Group

LETTERS

Direct electronic measurement of the spin Hall effect

S. O. Valenzuela[1]† & M. Tinkham[1]

The generation, manipulation and detection of spin-polarized electrons in nanostructures define the main challenges of spin-based electronics[1]. Among the different approaches for spin generation and manipulation, spin–orbit coupling—which couples the spin of an electron to its momentum—is attracting considerable interest. In a spin–orbit-coupled system, a non-zero spin current is predicted in a direction perpendicular to the applied electric field, giving rise to a spin Hall effect[2–4]. Consistent with this effect, electrically induced spin polarization was recently detected by optical techniques at the edges of a semiconductor channel[5] and in two-dimensional electron gases in semiconductor heterostructures[6,7]. Here we report electrical measurements of the spin Hall effect in a diffusive metallic conductor, using a ferromagnetic electrode in combination with a tunnel barrier to inject a spin-polarized current. In our devices, we observe an induced voltage that results exclusively from the conversion of the injected spin current into charge imbalance through the spin Hall effect. Such a voltage is proportional to the component of the injected spins that is perpendicular to the plane defined by the spin current direction and the voltage probes. These experiments reveal opportunities for efficient spin detection without the need for magnetic materials, which could lead to useful spintronics devices that integrate information processing and data storage.

The spin Hall effect (SHE), which was first described in refs 2 and 3 and more recently in ref. 4, was proposed to occur in paramagnetic materials as a consequence of the spin–orbit interaction. In analogy to the standard Hall effect, the SHE refers to the generation of a pure spin current transverse to an applied electric field that results in an accompanying spin imbalance in the system. These early theoretical studies considered an extrinsic[2–4,8] SHE originating from an asymmetric scattering for spin-up and spin-down electrons. It was pointed out that, after scattering off an impurity, there is a spin-dependent probability difference in the electron trajectories, which generates the spin imbalance. In the recently introduced intrinsic[9,10] SHE, spin imbalance is expected to occur even in the absence of scattering as a result of the band structure.

Several experimental schemes have been proposed to electrically detect the extrinsic SHE in metals[4,8,11,12]. However, these schemes are difficult to implement. Spin-related phenomena such as anisotropic magnetoresistance in ferromagnetic (FM) electrodes, spin dependent interface scattering, or standard and anomalous Hall effects could render the SHE signal unobservable. We use the measurement scheme in Fig. 1 to study the SHE isolated from such spurious phenomena.

It is natural to expect that, if a charge current induces a transverse spin imbalance through the spin–orbit interaction, a spin current will induce a transverse charge imbalance (and a measurable voltage) by the same mechanism[4]. According to the Onsager symmetry relations, these 'direct' and 'reverse' effects are mathematically equivalent and we simply refer to them as SHE. In our experiments (Fig. 1a), a ferromagnetic electrode (FM1) is used to inject spin-polarized electrons via a tunnel barrier in one of the arms of an aluminium (Al) Hall cross. As shown in Fig. 1b, the injected current I is driven away from the Hall cross, where only a pure spin current flows as a result of the spin injection. If spin–orbit scattering is present, the spin current could induce a transverse spin-Hall charge imbalance and generate a measurable voltage, V_{SH}. At a distance L_{SH} from the spin injector, the two transverse probes forming the Hall cross are used to measure such a voltage. The tunnel barrier[13] is important because it assures a uniformly distributed injection current, and enhances the polarization of the injected electrons[14,15]. The non-local nature of our measurement scheme eliminates spurious effects due to anisotropic magnetoresistance in the FM electrode or the anomalous and standard Hall effects. As no net charge current flows into the Hall cross, the direct generation of voltage by the standard Hall effect is precluded. Located at a distance L_{FM} from FM1 is a second ferromagnetic electrode (labelled FM2 in Fig. 1a) that is used for sample characterization, as explained below.

We prepare the devices with electron beam lithography and a two-angle shadow-mask evaporation technique, producing tunnel barriers *in situ*. The Al cross, with arms 400 nm and 60 nm wide, is first deposited at normal incidence onto a Si/SiO_2 substrate using electron beam evaporation. Next, the Al is oxidized in pure oxygen (150 mtorr for 40 min) to generate insulating Al_2O_3 barriers. After the vacuum is recovered, the two FM electrodes (400 nm and 250 nm wide, and 50 nm thick) are deposited under an angle of 50°, measured from the normal to the substrate surface. The FM electrodes form tunnel junctions where they overlap with the Al strip, with a typical tunnel resistance of 4 kΩ for FM1 and 6.5 kΩ for FM2. For the FM electrodes, we use CoFe (80 wt% Co), which provides a large polarization when combined with Al_2O_3 as a tunnelling barrier[14,16]. The difference in the widths of the FM electrodes is required to get different coercive fields. A crucial point in our fabrication procedure is that no image of the Hall cross is deposited during the evaporation of FM1 and FM2, as observed in the atomic force micrograph of Fig. 1a. The FM material deposits on the wall of the lithography mask and is removed by lift-off, preventing unwanted short circuits (see ref. 14 and Supplementary Fig. 1).

The spin polarization, P, of the electrons injected by FM1 depends on the effective tunnel conductance for spin-up and spin-down electrons, respectively $G_\uparrow$ and $G_\downarrow$, and can be written as $P = (G_\uparrow - G_\downarrow)/(G_\uparrow + G_\downarrow)$. The spin-polarized current causes unequal electrochemical potentials for the spin-up and spin-down populations in Al (Fig. 1c, top panel) and a spin current that flows to both sides of the contact and decays with the spin diffusion length, λ_{sf} (Fig. 1c, bottom panel). As a result, the charge imbalance that is built up at the edges of the Al strip and V_{SH} are expected to be proportional to P and to decay with λ_{sf}. Their magnitude is

[1]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. †Present address: Francis Bitter Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

© 2006 Nature Publishing Group

determined by the anomalous Hall operator[8,11,12,17], $\sigma_{SH}\ \hat{\boldsymbol{\sigma}} \times \mathbf{E}^{\sigma}$, where σ_{SH} denotes the spin Hall conductivity, σ is the spin index, and $\mathbf{E}^{\sigma}$ is an effective spin-dependent 'electric' field, which follows from the spin-dependent electrochemical potential μ^{σ} along the Al strip, that is, $\mathbf{E}^{\sigma}(\mathbf{r}) = -\nabla\mu^{\sigma}(\mathbf{r})$.

Spin imbalance in the Al film occurs with a defined spin direction given by the magnetization orientation of the FM electrode. Consequently, V_{SH} is expected to vary when a magnetic field perpendicular to the substrate, $B_{\perp}$, is applied and the magnetization $\mathbf{M}$ of the electrode is tilted out of the substrate plane. Defining θ as the angle between $\mathbf{M}$ and the electrode axis (Fig. 1a), we see, from the cross product in the anomalous Hall operator, that V_{SH} is proportional to $\sin\theta$, correlating with the component of $\mathbf{M}$ normal to the substrate. As discussed below, θ can be set to be any value between about $-90°$ and about $90°$, with $|B_{\perp}| < 2$ T.

We first characterize the properties of our device. FM1 and FM2, together with the Al strip, define a reference spin transistor[18–20] with a thin-film layout[13]. We use this spin transistor to obtain P, λ_{sf} and θ at $B_{\perp} \neq 0$. Measurements are performed at 4.2 K as represented in Fig. 1d and e (note the change in the direction of I in Fig. 1d and b). Both P and λ_{sf} are obtained by measuring the spin transresistance, $\Delta R = \Delta V/I$, as a function of L_{FM}, where ΔV is the difference in the output voltage between parallel and antiparallel magnetization configurations of the FM electrodes at zero magnetic field[13,18–22]. A lock-in amplifier is used with I equal to 50 μA. At larger currents, a marked decrease in the spin polarization is observed, as previously reported[15].

Figure 2b shows results for three different batches of samples with different Al thickness, t_{Al} (the experiment geometry is shown in Fig. 2a). As noted recently[14], in Al films λ_{sf} is strongly dependent on t_{Al}. We use this property to study the SHE in Al with different λ_{sf}. The top two sets of measurements correspond to $t_{Al} = 25$ nm (red circles and green diamonds), whereas the bottom one (blue squares) corresponds to $t_{Al} = 12$ nm. The data shown in red and green are almost identical, demonstrating the high reproducibility of our sample fabrication. By fitting the data to[13,18–21] $\Delta R = P^2 \frac{\lambda_{sf}}{\sigma_c A} \exp(-L_{FM}/\lambda_{sf})$, where σ_c is the charge conductivity of Al (σ_c(12 nm thickness) $= 1.05 \times 10^7\ \Omega^{-1}\,\mathrm{m}^{-1}$, σ_c(25 nm) $= 1.7 \times 10^7\ \Omega^{-1}\,\mathrm{m}^{-1}$) and A the cross-sectional area of the Al strip, we obtain $P = 0.28$, $\lambda_{sf}(12\,\mathrm{nm}) = 455 \pm 15$ nm and $\lambda_{sf}(25\,\mathrm{nm}) = 705 \pm 30$ nm.

The tilting angle can be obtained from the spin precession results shown in Fig. 2c, which are measured with $B_{\perp}$ between -3.5 T and 3.5 T. For small $B_{\perp}$, the magnetizations of the FM electrodes remain in plane owing to shape anisotropy, and the measurements show the Hanle effect associated with precessing spins. As $B_{\perp}$ increases, the magnetizations tilt out of plane. For large enough $B_{\perp}$, they orient completely along the field and the measurements saturate to a positive constant value. The output voltage normalized by its

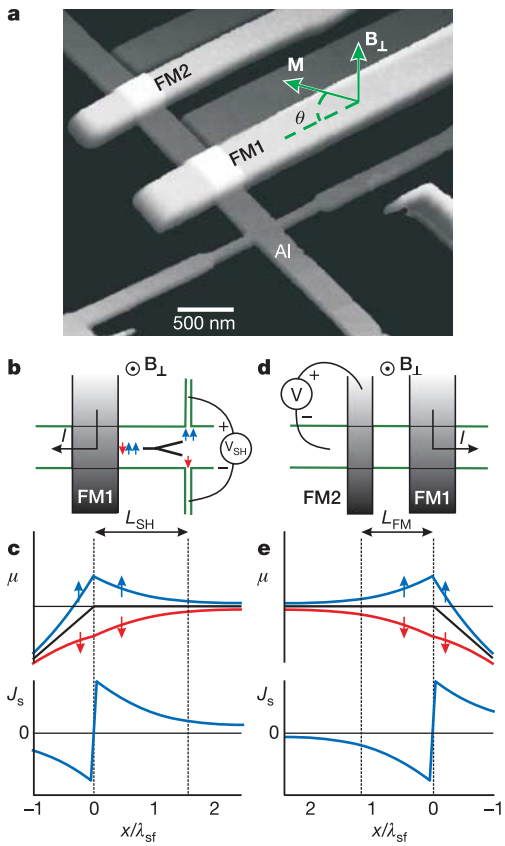


Figure 1 | Geometry of the devices and measurement schemes. a, Atomic force microscope image of a device. A thin aluminium (Al) Hall cross is oxidized and contacted with two ferromagnetic electrodes with different widths (FM1 and FM2). A magnetic field perpendicular to the substrate, $B_{\perp}$, sets the orientation of the magnetization M of FM1 (and FM2), which is characterized by an angle θ. **b**, Spin Hall measurement. A current I is injected out of FM1 into the Al film and away from the Hall cross. A spin Hall voltage, V_{SH}, is measured between the two Hall probes at a distance L_{SH} from the injection point. V_{SH} is caused by the separation of up and down spins due to spin–orbit interaction in combination with a pure spin current. **c**, Top: spatial dependence of the spin-up and spin-down electrochemical potentials, $\mu_{\uparrow,\downarrow}$. The black line represents the electrochemical potential of the electrons in the absence of spin injection. λ_{sf} is the spin diffusion length. Bottom: associated spin current, J_s. The polarized spins are injected near $x = 0$ and diffuse in both Al branches in opposite directions. The sign change in J_s reflects the flow direction. **d**, Spin-transistor measurement for device characterization. I is injected out of FM1 into the Al film and away from FM2, which is located at a distance L_{FM} from FM1. A voltage V is measured between FM2 and the left side of the Al film. **e**, As in **c** but for the conditions shown in **d**. Note that both V_{SH} in **b** and V in **d** vary with θ.

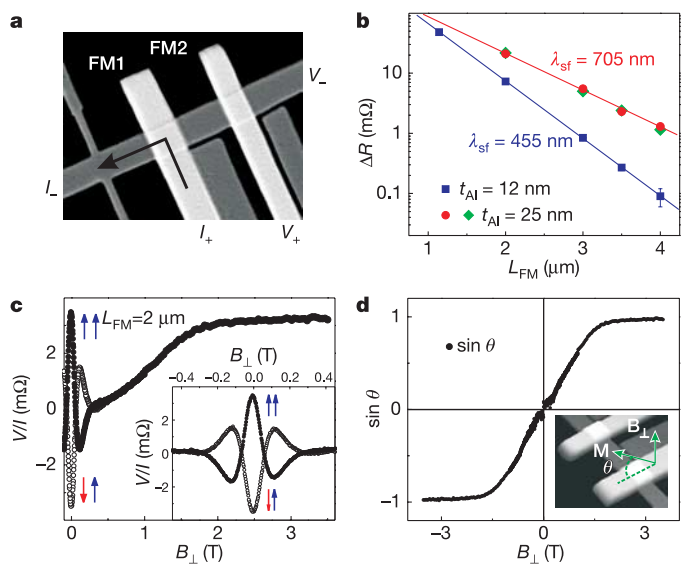


Figure 2 | Spin-transistor measurements and spin precession. a, Scanning electron micrograph of the device and the measurement scheme. **b**, $\Delta R = \Delta V/I$ as a function of L_{FM} for three sets of samples. Green diamonds and red circles are for an Al thickness $t_{Al} = 25$ nm, blue squares are for $t_{Al} = 12$ nm. The lines are best linear fits. λ_{sf}, spin diffusion length. For raw data, see Supplementary Fig. 2. The error bars are in most cases smaller than their data points, and represent instrumental errors and noise (standard deviation). **c**, Transresistance change due to spin precession as a function of $B_{\perp}$ (main panel, up to 3.5 T and inset, up to 0.5 T). Results are symmetric about $B_{\perp} = 0$. The arrows indicate the relative orientation of the magnetizations of FM1 and FM2. $t_{Al} = 12$ nm; $L_{FM} = 2$ μm. **d**, $\sin\theta$ as a function of $B_{\perp}$ extracted from data in **c**. Inset, magnetization direction of the FM electrodes relative to the substrate.

© 2006 Nature Publishing Group

value, V_0, at $B_\perp = 0$ has the form $V_\pm/V_0 = \pm f(B_\perp)\cos^2\theta + \sin^2\theta$, for initially parallel (+) and antiparallel (−) configurations of the magnetizations, with $f(B_\perp)$ a given function of $B_\perp$ (refs 13, 19, 20). By noting that $(V_+ + V_-)/V_0 = 2\sin^2\theta$, $f(B_\perp)$ is eliminated, and the calculation of $\sin\theta$ is straightforward. The result of this calculation is presented in Fig. 2d. At $B_\perp = 0$, $\theta = 0$ owing to shape anisotropy. When $B_\perp$ is applied, the magnetization follows the Stoner-Wohlfarth model[23] with a saturation field $B_\perp^{sat}$ of about 1.55 T for which $\sin\theta$ approaches one, and the magnetization aligns with the field.

Having determined P, λ_{sf} and θ, we study the SHE using the measurement configuration shown in Fig. 1b. A change in $B_\perp$ results in a change in V_{SH}. By considering the spin diffusion using a semiclassical Boltzmann approximation, the spin Hall resistance $R_{SH} = V_{SH}/I$ at L_{SH} (see Fig. 1b) can be calculated[8,24]:

$$R_{SH} = \frac{\Delta R_{SH}}{2}\sin\theta \qquad (1)$$

with

$$\Delta R_{SH} = \frac{P}{t_{Al}}\frac{\sigma_{SH}}{\sigma_c^2}\exp(-L_{SH}/\lambda_{sf}) \qquad (2)$$

Here it is assumed that the distance between the two voltage probes (Al strip width) is smaller than λ_{sf}. ΔR_{SH} in equation (2) was obtained as in refs 8 and 24. The factors $\sin\theta$ and 1/2 in equation (1) account for the orientation of the spins and the fact that they diffuse away from FM1 into two (Al) branches instead of one, as in ref. 24. Note that, in the present configuration, the measurements are not affected by spin precession because the component of the spins perpendicular to the substrate (or the transverse component of the anomalous velocity) is not modified by this effect.

Figure 3 shows R_{SH} as a function of $B_\perp$ for samples with $L_{SH} = 860$, 590 and 480 nm (circles) and $t_{Al} = 12$ nm. The sign convention is shown in the top inset, where a positive $B_\perp$ is pointing out of the page. $B_\perp$ is swept between −3.5 T and 3.5 T. We observe a linear response around $B_\perp = 0$, followed by a saturation on the scale of $B_\perp^{sat}$, both for positive and negative $B_\perp$. Similar results were obtained in about 20 samples with different t_{Al}.

The saturation in R_{SH} for $|B_\perp| > B_\perp^{sat}$ strongly suggests that the device output is related to the magnetization orientation of the FM electrode (Fig. 2d) and the SHE. This idea is further reinforced by comparing $R_{SH}(B_\perp)$ with the magnetization component perpendicular to the substrate, which as discussed above is proportional to $\sin\theta(B_\perp)$ (lines in Fig. 3). The agreement is excellent; the proportionality between R_{SH} and $\sin\theta$ in equation (1) is closely followed.

For L_{SH} smaller than λ_{sf}, a small linear term in $R_{SH}(B_\perp)$ becomes apparent (see bottom panel of Fig. 3 and its inset). We believe that this term arises from the contribution of the orbital Lorentz force to the spin-Hall induced charge current[12], which is a second-order effect that decays faster with L_{SH} than the spin Hall voltage itself.

Measurements were again performed with I equal to 50 μA. Larger currents lead to a decrease in R_{SH} as the voltage drop at the tunnel junction increases. This correlates with the decrease in the spin polarization observed with the spin-transistor measurements[15], and further supports the interpretation of our results as being spin-related.

Consistent with equation (2), the overall change of R_{SH}, ΔR_{SH}, decreases as a function of L_{SH} (Fig. 3). By fitting the magnetic field dependence of R_{SH} in Fig. 3 to equation (1), we have obtained ΔR_{SH} for each sample. The results are shown in Fig. 4 as a function of L_{SH}, which are then fitted to equation (2) in order to obtain λ_{sf} and σ_{SH}. We find $\lambda_{sf}(12\,\text{nm}) = 490 \pm 80$ nm, and $\lambda_{sf}(25\,\text{nm}) = 735 \pm 130$ nm in good agreement with the values obtained independently with the control spin transistor (Fig. 2b). By using $P = 0.28$, we find $\sigma_{SH}(12\,\text{nm}) = (3.4 \pm 0.6) \times 10^3\,\Omega^{-1}\,\text{m}^{-1}$ and $\sigma_{SH}(25\,\text{nm}) = (2.7 \pm 0.6) \times 10^3\,\Omega^{-1}\,\text{m}^{-1}$. The ratio $\sigma_{SH}/\sigma_c \approx (1\text{–}3) \times 10^{-4}$ is of the same order of magnitude as the one obtained experimentally[5] and theoretically[25,26] for GaAs. The predicted σ_{SH}, when considering short-range

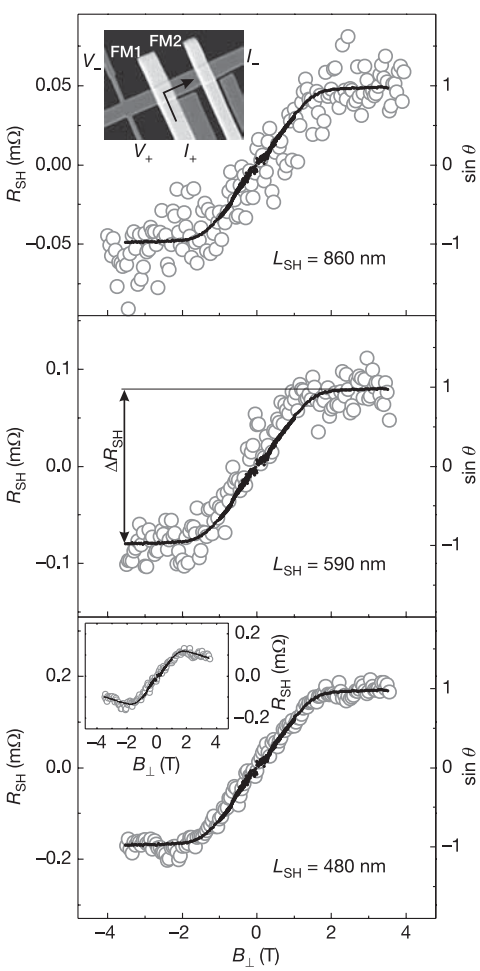


Figure 3 | Spin Hall effect. Spin Hall resistance R_{SH} versus perpendicular field $B_\perp$, for $t_{Al} = 12$ nm and $L_{SH} = 860$ nm (top), $L_{SH} = 590$ nm (middle) and $L_{SH} = 480$ nm (bottom). Top panel inset, a scanning electron micrograph of the device and the measurement scheme with the sign convention. Top and middle panels and inset in the bottom panel show raw data. In the bottom panel, the linear background was subtracted from the data in the inset. For comparison, the measured value of $\sin\theta$ is shown in each panel (black line). Note the decrease in ΔR_{SH} with L_{SH}.

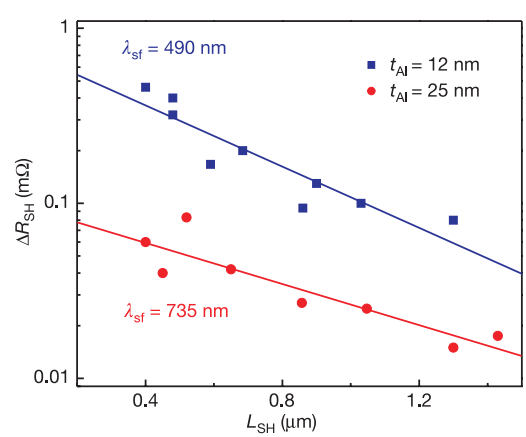


Figure 4 | Overall change of the spin Hall resistance, ΔR_{SH}, between large negative and large positive $B_\perp$. ΔR_{SH} is obtained by a best fit to equation (1) using data in Fig. 3. Solid lines are best fits to equation (2). Blue squares (red circles) are for aluminium thickness $t_{Al} = 12$ (25) nm. λ_{sf}, spin diffusion length.

© 2006 Nature Publishing Group

(δ-function) scattering centres, is $\alpha\hbar e^2 N_0/3m$ (refs 8, 11), where $\hbar$ is Planck's constant divided by 2π, $N_0 = 2.4 \times 10^{28}$ states $\mathrm{eV}^{-1}\,\mathrm{m}^{-3}$ is the density of states of Al at the Fermi energy[27], $\alpha \approx 0.006$ is the dimensionless spin–orbit coupling constant of Al[12], and e and m are the charge and mass of the electron. Without free parameters, we obtain $\sigma_{\mathrm{SH}} \approx 10^3\,\Omega^{-1}\,\mathrm{m}^{-1}$ and a ratio $\sigma_{\mathrm{SH}}/\sigma_{\mathrm{c}} \approx (0.4\text{–}1) \times 10^{-4}$, in reasonable agreement with the experimental results.

We have thus presented electrical measurements of the SHE in a diffusive metallic conductor (aluminium), and have obtained the spin-Hall conductivity. We have studied spin injection and spin relaxation with both reference spin transistors and spin Hall crosses, obtaining consistent results.

This work demonstrates new means to study spin-related phenomena. It shows that the SHE can be used to directly measure spin-polarized currents without the need of ferromagnets. Spin precession does not modify the spin Hall voltage for magnetic fields perpendicular to the substrate, as in the configuration discussed here. However, preliminary measurements (S.O.V. and M.T., unpublished data) with a magnetic field applied parallel to the Al strip suggest the generation of a spin Hall voltage that is a consequence of out-of-plane spins induced by spin precession.

Received 10 January; accepted 23 May 2006.

1. Wolf, S. A. *et al.* Spintronics: a spin-based electronics vision for the future. *Science* **294,** 1488–1495 (2001).
2. Dyakonov, M. I. & Perel, V. I. Possibility of orienting electron spins with current. *JETP Lett.* **13,** 467–469 (1971).
3. Dyakonov, M. I. & Perel, V. I. Current-induced spin orientation of electrons in semiconductors. *Phys. Lett. A* **35,** 459–460 (1971).
4. Hirsch, J. E. Spin Hall effect. *Phys. Rev. Lett.* **83,** 1834–1837 (1999).
5. Kato, Y. K., Myers, R. C., Gossard, A. C. & Awschalom, D. D. Observation of the spin Hall effect in semiconductors. *Science* **306,** 1910–1913 (2004).
6. Wunderlich, J., Kaestner, B., Sinova, J. & Jungwirth, T. Experimental observation of the spin Hall effect in a two-dimensional spin-orbit coupled semiconductor system. *Phys. Rev. Lett.* **94,** 047204 (2005).
7. Sih, V. *et al.* Spatial imaging of the spin Hall effect and current-induced polarization in two-dimensional electron gases. *Nature Phys.* **1,** 31–35 (2005).
8. Zhang, S. Spin Hall effect in the presence of spin diffusion. *Phys. Rev. Lett.* **85,** 393–396 (2000).
9. Murakami, S., Nagaosa, N. & Zhang, S. C. Dissipationless quantum spin current at room temperature. *Science* **301,** 1348–1351 (2003).
10. Sinova, J. *et al.* Universal intrinsic spin Hall effect. *Phys. Rev. Lett.* **92,** 126603 (2004).
11. Shchelushkin, R. V. & Brataas, A. Spin Hall effects in diffusive normal metals. *Phys. Rev. B* **71,** 045123 (2005).
12. Shchelushkin, R. V. & Brataas, A. Spin Hall effect, Hall effect, and spin precession in diffusive normal metals. *Phys. Rev. B* **72,** 073110 (2005).
13. Jedema, F. J., Heersche, H. B., Filip, A. T., Baselmans, J. J. A. & van Wees, B. J. Electrical detection of spin precession in a metallic spin valve. *Nature* **416,** 713–716 (2002).
14. Valenzuela, S. O. & Tinkham, M. Spin-polarized tunneling in room-temperature spin valves. *Appl. Phys. Lett.* **85,** 5914–5916 (2004).
15. Valenzuela, S. O., Monsma, D. J., Marcus, C. M., Narayanamurti, V. & Tinkham, M. Spin polarized tunneling at finite bias. *Phys. Rev. Lett.* **94,** 196601 (2005).
16. Monsma, D. J. & Parkin, S. S. P. Spin polarization of tunneling current from ferromagnet/Al_2O_3 interfaces using copper-doped aluminum superconducting films. *Appl. Phys. Lett.* **77,** 720–722 (2000).
17. Qi, Y. & Zhang, S. Spin diffusion at finite electric and magnetic fields. *Phys. Rev. B* **67,** 052407 (2003).
18. Johnson, M. & Silsbee, R. H. Interfacial charge-spin coupling: injection and detection of spin magnetization in metals. *Phys. Rev. Lett.* **55,** 1790–1793 (1985).
19. Johnson, M. & Silsbee, R. H. Coupling of electronic charge and spin at a ferromagnetic-paramagnetic metal interface. *Phys. Rev. B* **37,** 5312–5325 (1988).
20. Johnson, M. & Silsbee, R. H. Spin injection experiment. *Phys. Rev. B* **37,** 5326–5335 (1988).
21. Ji, Y., Hoffmann, A., Jiang, J. S. & Bader, S. D. Spin injection, diffusion, and detection in lateral spin-valves. *Appl. Phys. Lett.* **85,** 6218–6220 (2004).
22. Urech, M., Korenivski, V., Poli, N. & Haviland, D. B. Direct demonstration of decoupling of spin and charge currents in nanostructures. *Nano Lett.* **6,** 871–874 (2006).
23. O'Handley, R. C. *Modern Magnetic Materials* (Wiley & Sons, New York, 2000).
24. Zhang, S. Probing spin currents in semiconductors. *J. Appl. Phys.* **89,** 7564–7566 (2001).
25. Engel, H. A., Halperin, B. I. & Rashba, E. I. Theory of spin Hall conductivity in *n*-doped GaAs. *Phys. Rev. Lett.* **95,** 166605 (2005).
26. Tse, W. K. & Das Sarma, S. Spin Hall effect in doped semiconductor structures. *Phys. Rev. Lett.* **96,** 056601 (2006).
27. Papaconstantopoulos, D. A. *Handbook of the Band Structure of Elemental Solids* (Plenum, New York, 1986).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. DiCarlo, H. A. Engel, D. J. Monsma and W. D. Oliver for a critical reading of the manuscript. This research was supported in part by the US National Science Foundation and the US Office of Naval Research.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.O.V. (sov@mit.edu).

© 2006 Nature Publishing Group

LETTERS

Ultrasensitive solution-cast quantum dot photodetectors

Gerasimos Konstantatos[1], Ian Howard[1], Armin Fischer[1], Sjoerd Hoogland[1], Jason Clifford[1], Ethan Klem[1], Larissa Levina[1] & Edward H. Sargent[1]

Solution-processed electronic[1] and optoelectronic[2–5] devices offer low cost, large device area, physical flexibility and convenient materials integration compared to conventional epitaxially grown, lattice-matched, crystalline semiconductor devices. Although the electronic or optoelectronic performance of these solution-processed devices is typically inferior to that of those fabricated by conventional routes, this can be tolerated for some applications in view of the other benefits. Here we report the fabrication of solution-processed infrared photodetectors that are superior in their normalized detectivity (D^*, the figure of merit for detector sensitivity) to the best epitaxially grown devices operating at room temperature. We produced the devices in a single solution-processing step, overcoating a prefabricated planar electrode array with an unpatterned layer of PbS colloidal quantum dot nanocrystals. The devices showed large photoconductive gains with responsivities greater than 10^3 A W^{-1}. The best devices exhibited a normalized detectivity D^* of 1.8×10^{13} jones (1 jones = 1 cm Hz$^{1/2}$ W^{-1}) at 1.3 μm at room temperature: today's highest performance infrared photodetectors are photovoltaic devices made from epitaxially grown InGaAs that exhibit peak D^* in the 10^{12} jones range at room temperature, whereas the previous record for D^* from a photoconductive detector lies at 10^{11} jones. The tailored selection of absorption onset energy through the quantum size effect, combined with deliberate engineering of the sequence of nanoparticle fusing and surface trap functionalization, underlie the superior performance achieved in this readily fabricated family of devices.

Sensing beyond 1 μm into the short-wavelength infrared is critical to environmental monitoring and remote sensing[6–9], fibre-optic communications, night-time surveillance[10] and emerging medical imaging modalities[11–13]. The photosensitive properties of silicon deteriorate rapidly beyond 800 nm and end abruptly at 1.1 μm. It is therefore of great interest to produce infrared photodetectors via facile processing from solution, thereby enabling convenient monolithic integration with silicon integrated circuits.

Infrared photodetector sensitivity is quantified using the parameter D*, the normalized detectivity measured in units of jones (cm Hz$^{1/2}$ W^{-1}). D^* is given as $(A\Delta f)^{1/2}R/i_n$, where A is the effective area of the detector in cm^2, Δf the electrical bandwidth in Hz, and R the responsivity in A W^{-1} measured under the same conditions as the noise current i_n in A. The material figure of merit D^* allows comparison among devices of different areas and geometries. The device figure of merit, noise equivalent power (NEP)—the minimum impinging optical power that a detector can distinguish from noise—is related to D^* by NEP $= (A\Delta f)^{1/2}/D^*$.

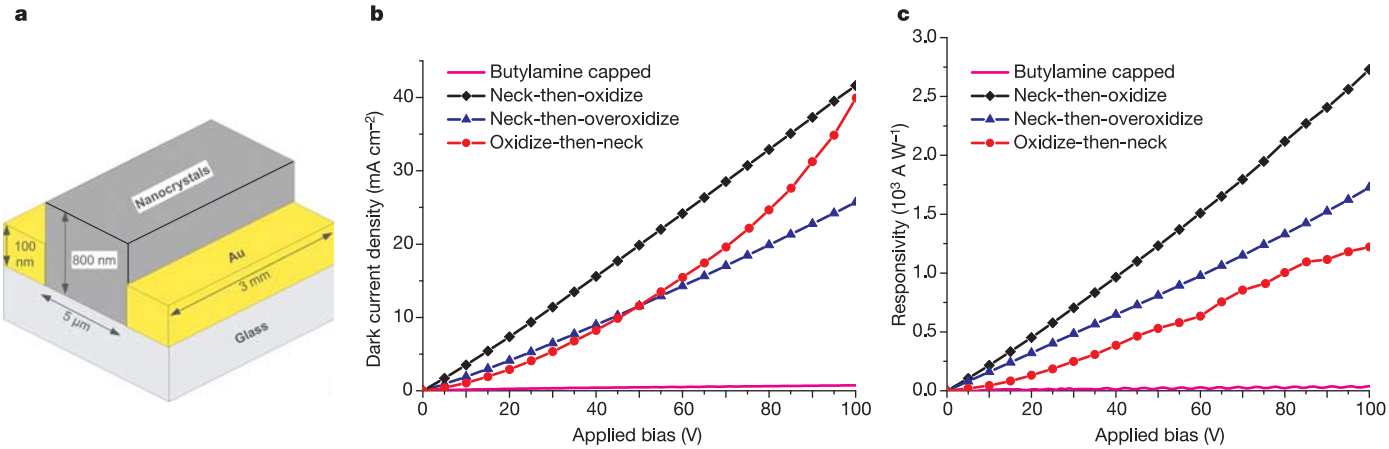


Figure 1 | Device structure and electro-optic characteristics for device classes investigated. **a**, Device structure. Smooth films of controlled thickness were formed by spin-coating nanocrystals from chloroform solution. **b**, Dark current density. Nanocrystal necking by way of soaking in methanol resulted in two orders of magnitude increase in dark current density. Devices made using nanocrystals exposed to oxygen before film formation ('oxidize-then-neck') show a superlinear I–V behaviour characteristic of field-assisted transport. Devices made using nanocrystals necked before oxidization exhibit linear (field-independent) behaviour. Further oxidation of neck-then-oxidize devices (neck-then-overoxidize) leads to a decrease of conductivity owing to excessive oxide formation. **c**, Responsivity as a function of applied bias. Necked nanocrystal devices show comparable responsivities, consistent with similar carrier mobilities and trap state lifetimes.

[1]Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4 Canada.

© 2006 Nature Publishing Group

The thermodynamic upper bound on D^* is fundamentally determined by the bandgap of the semiconductor used as the light-absorbing medium. Smaller-bandgap materials absorb more low-energy thermal photons, increasing noise. The ideal medium for photodetection would therefore be one that is spectrally tunable, tailored optimally to use the largest effective bandgap that nevertheless absorbs all photons of interest in a particular application. Practical limits on experimentally achieved D^* values are explained in terms of excess noise. Polycrystalline photoconductive detectors[14], for example, exhibit multiplicative noise[15] associated with the product of transport noise with fundamental generation-recombination noise in band-edge and trap states.

Colloidal semiconductor quantum dots offer a simultaneous solution to the aforementioned challenges: they are processed from solution, allowing convenient integration with any substrate; they can be tuned by way of the quantum size effect[16], allowing the effective bandgap to be selected optimally for a given application; and their transport and trap state properties can be separately controlled through the engineering of ligands and the oxidation of nanoparticle surfaces either before or after film formation. Cross-linking of semiconductor nanoparticles has previously been shown to result in marked increases in film conductivity[17,18]. In addition, visible photoconduction has been reported in pure CdSe nanocrystal films following chemical treatment[19–21]. However, there exists only one previous observation of infrared photoconductivity[5], and there have been no previous reports of quantified photodetector sensitivity from such materials.

We fabricated photoconductive detectors by spin-coating colloidal quantum dots from solution onto gold interdigitated electrodes. The inter-electrode separation was 5 μm and the metal electrode height was 100 nm. The resultant structure is depicted in Fig. 1a. The thickness of the quantum dot films formed was 800 nm.

Devices synthesized using PbS nanocrystals capped with oleic acid[22] were found to be insulating. The ~2.5-nm-long ligand inhibited carrier transport among the nanocrystals. To improve carrier mobility, we used a post-synthetic ligand exchange to replace oleic acid with *n*-butylamine of length ~0.6 nm. Ligand exchange, confirmed via Fourier transform infrared spectroscopy (FTIR; Supplementary Information 1), resulted in decreased interparticle spacing as observed via transmission electron microscopy (TEM; Supplementary Information 1). The conductance after exchange was low but measurable, producing a dark current density of ~0.7 mA cm^{-2} at 100 V bias. We therefore immersed devices in methanol (MeOH) for 2 hours to remove butylamine ligands and further decrease interparticle separation, causing necking, or cross-linking, at points where adjacent nanoparticles' surfaces are in contact with one another (Supplementary Information 2). The dark current density increased by two orders of magnitude to ~40 mA cm^{-2}.

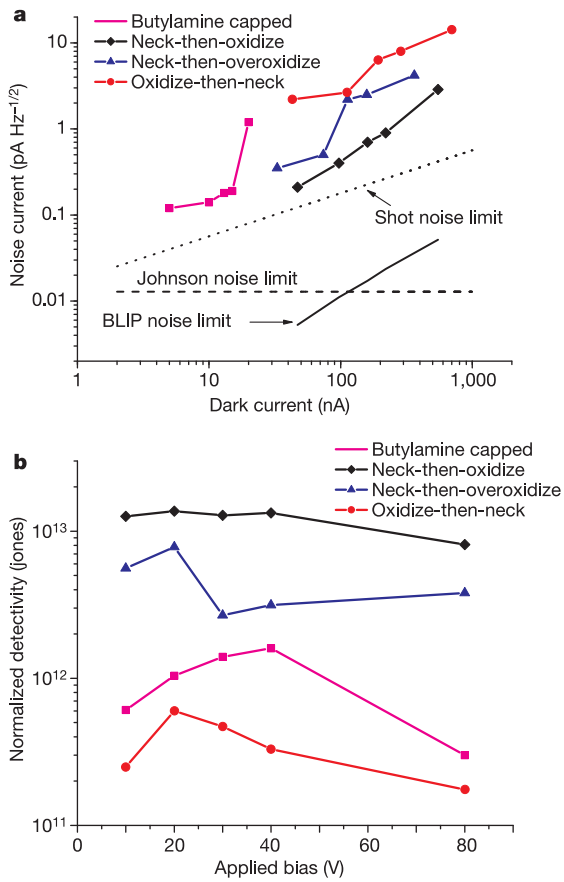


Figure 2 | Noise characteristics and resultant normalized detectivity of the different device classes investigated. a, Measured noise current as a function of measured dark current. Neck-then-oxidize devices exhibited the lowest noise current that approaches within 3 dB the shot noise limit. Oxidize-then-neck nanocrystal devices had the highest noise current consistent with multiplicative noise. Devices formed from neck-then-overoxidize nanocrystals showed lower noise levels than the oxidize-then-neck nanocrystal devices although they contained larger amounts of oxide. This indicates the criticality of the oxidation step in the fabrication process. The Johnson noise limit, the shot-noise limit, and fundamental background-limited thermodynamic (BLIP) noise current of the best-performing devices (neck-then-oxidize) are also plotted for comparison. **b**, Normalized detectivity D^* as a function of applied bias. Oxidize-then-neck devices demonstrated the lowest detectivity. The highest detectivity was obtained from neck-then-oxidize devices measured at a modulation frequency of 30 Hz, and reached 1.3×10^{13} jones at 975 nm excitation wavelength.

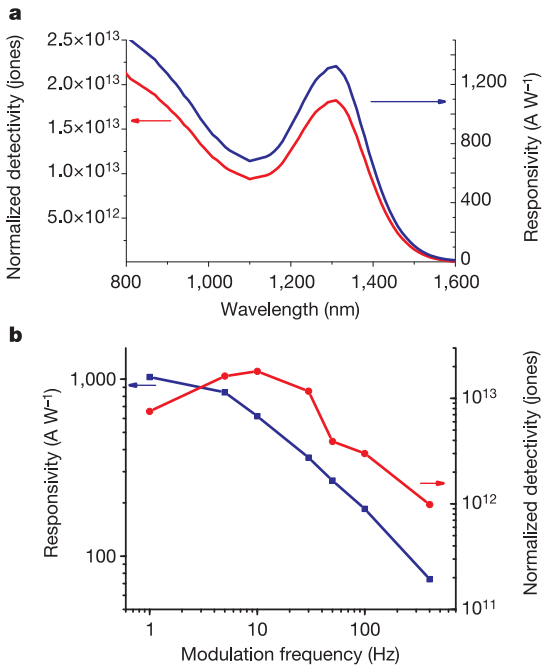


Figure 3 | Photodetector performance characteristics of the highest-sensitivity class of devices. a, Spectra of responsivity and normalized detectivity D^* of neck-then-oxidize nanocrystal devices. The applied bias is 40 V and the electrical frequency 10 Hz. D^* was measured to be 1.8×10^{13} jones at the excitonic peak wavelength. **b**, Electrical frequency response of the same devices under 40 V bias. The 3-dB bandwidth of the detectors is ~18 Hz, consistent with the longest excited-state carrier lifetime in these devices. High sensitivity ($D^* > 10^{13}$ jones) is retained at imaging rates of 30 frames per second.

© 2006 Nature Publishing Group

Photoconductivity relies on long-lived trap states to enable photoconductive gain. However, to suppress the multiplicative noise to which conventional polycrystalline photoconductive detectors are prone, it is necessary to minimize transport noise associated with trap states lying at the interface between quantum dots. We therefore fabricated two classes of devices, one intended to minimize multiplicative noise, the other to serve as control. In devices labelled 'neck-then-oxidize', we maintained oxygen-free conditions throughout all process steps before film fabrication, sensitizing by means of oxidation only at the end. In devices labelled 'oxidize-then-neck', we precipitated the colloidal quantum dots from their butylamine solvent in oxygen-rich ambient conditions.

X-ray photoelectron spectroscopy (XPS; Supplementary Information 3) showed that butylamine-capped nanocrystals precipitated in inert conditions were free from $PbSO_4$, the trap-state-forming oxide native to PbS. Butylamine-capped nanocrystals precipitated in inert conditions and then treated by methanol soaking (neck-then-oxidize) consisted predominantly of PbS, but exhibited a measurable $PbSO_4$ level estimated at 5% (that is, 5% of Pb atoms were bound to sulphate groups instead of sulphur). Butylamine-capped nanocrystals precipitated in an air ambient (oxidize-then-neck) exhibited a significant amount of $PbSO_4$ formation, with a sulphate fraction now at 25%. These observations regarding sulphate levels were confirmed using FTIR (Supplementary Information 3).

As shown in Fig. 1b, the order of operations in necking versus oxidizing does not significantly influence the magnitude of the dark current. However, it does influence the shape of the dark current versus voltage (I–V) characteristic. Neck-then-oxidize devices exhibit a linear (field-independent) I–V characteristic, suggesting that nanocrystal necking was achieved and the dominant conduction mechanism was variable range hopping[23]. Oxidize-then-neck devices exhibited a superlinear (higher slope at higher biases) I–V characteristic, suggesting that transport in these devices is field-assisted owing to sulphate barriers formed along the pathway of carrier transport.

The signal component of the sensitivity of photodetectors is quantified by the responsivity. Figure 1c shows the measured responsivity as a function of the applied bias for devices excited using a 975-nm laser. The optical power impinging on each device was ~80 pW (Supplementary Information 4). Responsivity correlated closely with dark current, indicating that high responsivities were associated with the successful removal of ligands that led to nanocrystal necking. Long trap state lifetimes, also necessary for photoconductive gain, were amply achieved in all devices with the aid of oxidation of the high-surface-to-volume nanometre-sized semiconductor particles.

We investigated experimentally the transit time and the longest tail of carrier lifetime to provide independent confirmation of the existence of such strikingly high photoconductive gain. The response of the detector to a 7-ns optical pulse was found to persist over tens of milliseconds, attributable to the longest-lived population of trap states introduced by oxidation[24] (Supplementary Information 5). The response exhibits multiple lifetime components that extend from microseconds to several milliseconds (decay components were found at ~20 μs, ~200 μs, 2 ms, 7 ms and 70 ms). We obtained transit times of 500 ns at a bias of 100 V (ref. 25), finding that transit time depended linearly on bias with a slope corresponding to a mobility of 0.1 $cm^2 V^{-1} s^{-1}$, consistent with prior reports[1]. The ratio of the longest component of carrier lifetime to the transit time was thus of the order of 10,000. Our observed responsivities of 2,700 $A W^{-1}$ are thus explainable by photoconductive gain, given our films' absorbance of 0.3 at an optical wavelength of 975 nm. We note also that this high responsivity was observed under the low-level optical power conditions relevant to ultrasensitive detection. As we increased illumination intensity, the longest-lived trap states became filled, and shorter-lived, hence lower-gain trap states began to account for a significant component of carrier lifetime. The devices were thus most sensitive under low-light illumination conditions, and exhibit intrinsic dynamic-range-enhancing gain compression under increasing illumination.

Achieving sensitive detection requires not only high responsivity, but also low noise. We plot in Fig. 2a the measured noise current versus the dark current for each class of devices. Neck-then-oxidize nanoparticles approach the shot noise limit within 3 dB. Oxidize-then-neck devices lie at least 15 dB from fundamental noise limits. We measured the noise current and, combining this with the responsivity, obtained the normalized detectivity D^* (Supplementary Information 6). Figure 2b shows D^* of the photodetectors at different bias conditions, measured at an optical excitation wavelength of 975 nm and a modulation frequency of 30 Hz. The most sensitive devices were those labelled neck-then-oxidize, from which $D^* > 10^{13}$ jones was obtained.

The electro-optic characteristics of a typical ultrasensitive detector are presented in Fig. 3 (Supplementary Information 7). Figure 3a shows the spectral responsivity and normalized detectivity for electrical modulation frequency 30 Hz and applied bias 40 V. The responsivity spectrum corresponds closely to the absorption spectrum. Figure 3b presents the frequency response of the detector under modulated illumination. The 3-dB bandwidth is ~18 Hz, consistent with the longest carrier lifetime component at 70 ms. D^* remains in excess of 10^{13} jones at the 30 frames-per-second required for imaging.

Reproducibility of all results reported here was confirmed by fabricating several devices using nanocrystals from several different synthesis batches. Preliminary studies of device lifetime indicate that the performance of devices stored in air for 2 weeks does not change by more than 20%; and that devices stored in a nitrogen glovebox do not, over the course of months, change more than 20% in performance.

Received 21 December 2005; accepted 28 April 2006.

1. Talapin, D. & Murray, C. B. PbSe nanocrystal solids for n- and p-channel thin film field-effect transistors. *Science* **310,** 86–89 (2005).
2. Tessler, N., Medvedev, V., Kazes, M., Kan, S. & Banin, U. Efficient near-infrared polymer nanocrystal light-emitting diodes. *Science* **295,** 1506–1508 (2002).
3. Konstantatos, G., Huang, C., Levina, L., Lu, Z. & Sargent, E. H. Efficient infrared electroluminescent devices using solution-processed colloidal quantum dots. *Adv. Funct. Mater.* **15,** 1865–1869 (2005).
4. Hoogland, S. *et al.* A solution-processed 1.53 μm quantum dot laser with temperature-invariant emission wavelength. *Opt. Express* **14,** 3273–3281 (2006).
5. McDonald, S. A. *et al.* Solution-processed PbS quantum dot infrared photodetectors and photovoltaics. *Nature Mater.* **4,** 138–142 (2005).
6. Schödel, R. *et al.* A star in a 15.2-year orbit around the supermassive black hole at the centre of the Milky Way. *Nature* **419,** 694–696 (2002).
7. Rawlings, S. *et al.* A radio galaxy at redshift 4.41. *Nature* **383,** 502–505 (1996).
8. Ettl, R., Chao, I., Diederich, F. & Whetten, R. L. Isolation of C_{76}, a chiral (D_2) allotrope of carbon. *Nature* **353,** 149–153 (1991).
9. Walker, J. Tunable alkali halide lasers. *Nature* **256,** 695 (1975).
10. Ettenberg, M. A little night vision. *Adv. Imaging* **20,** 29–32 (2005).
11. Sargent, E. H. Infrared quantum dots. *Adv. Mater.* **17,** 515–522 (2005).
12. Lim1, Y. T. *et al.* Selection of quantum dot wavelengths for biomedical assays and imaging. *Mol. Imaging* **2,** 50–64 (2003).
13. Kim, S. *et al.* Near-infrared fluorescent type II quantum dots for sentinel lymph node mapping. *Nature Biotechnol.* **22,** 93–97 (2004).
14. Petritz, R. L. Theory of photoconductivity in semiconductor films. *Phys. Rev. B* **104,** 1508–1516 (1956).
15. Carbone, A. & Mazzetti, P. Grain-boundary effects on photocurrent fluctuations in polycrystalline photoconductors. *Phys. Rev. B* **57,** 2454–2460 (1998).
16. Murray, C. B., Norris, D. J. & Bawendi, M. G. Synthesis and characterization of nearly monodisperse CdE (E = sulfur, selenium, tellurium) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115,** 8706–8715 (1993).
17. Yu, D., Wang, C. & Guyot-Sionnest, P. n-Type conducting CdSe nanocrystal solids. *Science* **300,** 1277–1280 (2003).
18. Wessels, J. M. *et al.* Optical and electrical properties of three-dimensional interlinked gold nanoparticle assemblies. *J. Am. Chem. Soc.* **126,** 3349–3356 (2004).
19. Leatherdale, C. A. *et al.* Photoconductivity in CdSe quantum dot solids. *Phys. Rev. B* **62,** 2669–2680 (2000).
20. Jarosz, M. V., Porter, V. J., Fisher, B. R., Kastner, M. A. & Bawendi, M. G. Photoconductivity of treated CdSe quantum dot films exhibiting increased exciton ionization efficiency. *Phys. Rev. B* **70,** 195327 (2004).

© 2006 Nature Publishing Group

21. Oertel, D. C., Bawendi, M. G., Arango, A. C. & Bulovic, V. Photodetectors based on treated CdSe quantum-dot films. *Appl. Phys. Lett.* **87,** 213505 (2005).
22. Hines, M. A. & Scholes, G. D. Colloidal PbS nanocrystals with size-tunable near-infrared emission: observation of post-synthesis self-narrowing of the particle size distribution. *Adv. Mater.* **15,** 1844–1849 (2003).
23. Yu, D., Wang, C., Wehrenberg, B. L. & Guyot-Sionnest, P. Variable range hopping mechanism in semiconductor nanocrystal solids. *Phys. Rev. Lett.* **92,** 216802 (2004).
24. Rahada, R. H. & Minden, H. T. Photosensitization of PbS films. *Phys. Rev.* **102,** 1258–1262 (1956).
25. Scher, H. Anomalous transit-time dispersion in amorphous solids. *Phys. Rev. B* **12,** 2455–2477 (1975).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Grozea and Z. H. Lu for XPS, E. Istrate for discussions, and V. Sukhovatkin for assistance in optical measurements. This research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Canada Foundation for Innovation, the Province of Ontario, and the Canada Research Chairs programme.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.H.S. (ted.sargent@utoronto.ca).

© 2006 Nature Publishing Group

LETTERS

Strong correlation between levels of tropospheric hydroxyl radicals and solar ultraviolet radiation

Franz Rohrer[1] & Harald Berresheim[2]

The most important chemical cleaning agent of the atmosphere is the hydroxyl radical[1,2], OH. It determines the oxidizing power of the atmosphere, and thereby controls the removal of nearly all gaseous atmospheric pollutants[3,4]. The atmospheric supply of OH is limited, however, and could be overcome by consumption due to increasing pollution and climate change[4–6], with detrimental feedback effects. To date, the high variability of OH concentrations has prevented the use of local observations to monitor possible trends in the concentration of this species. Here we present and analyse long-term measurements of atmospheric OH concentrations, which were taken between 1999 and 2003 at the Meteorological Observatory Hohenpeissenberg in southern Germany. We find that the concentration of OH can be described by a surprisingly linear dependence on solar ultraviolet radiation throughout the measurement period, despite the fact that OH concentrations are influenced by thousands of reactants. A detailed numerical model of atmospheric reactions and measured trace gas concentrations indicates that the observed correlation results from compensations between individual processes affecting OH, but that a full understanding of these interactions may not be possible on the basis of our current knowledge of atmospheric chemistry. As a consequence of the stable relationship between OH concentrations and ultraviolet radiation that we observe, we infer that there is no long-term trend in the level of OH in the Hohenpeissenberg data set.

The central role of OH in tropospheric chemistry was already recognized and described in 1971[1]. A major pathway for production of OH radicals is the photolysis of ozone by solar UV-B. This initial photolytic process is described by the photolysis frequency of ozone, $J(O^1D)$, the rate coefficient of the first reaction below. It generates excited $O(^1D)$ atoms, which are precursors for OH:

$$O_3 + h\nu \rightarrow O(^1D) + O_2 \ (\lambda < 330\,\mathrm{nm})$$

$$O(^1D) + H_2O \rightarrow 2OH$$

Once formed, OH radicals react with tropospheric trace constituents within about 1 second. Owing to this short chemical lifetime, ambient OH concentrations rarely exceed 10^7 molecules cm^{-3} and are difficult to measure. Reliable measurements have been possible only since the early 1990s[7–14]. Here we present the first (to our knowledge) long-term data set of OH recorded over 5 years (April 1999–December 2003). It was measured at the Meteorological Observatory Hohenpeissenberg (MOHp), a Global Atmosphere Watch (GAW[15,16]) site in rural southern Germany. Figure 1 shows the data in relation to the concurrently measured $J(O^1D)$ levels. In view of the variety of meteorological and chemical conditions of the air masses in the course of five years (Supplementary Table 1), we expected to find significant dependencies of OH on chemical composition changes. Surprisingly, our results show a linear and compact relation between OH and $J(O^1D)$ over the entire period. Essentially this means that the whole OH data set can be described by a single variable, $J(O^1D)$, with a linear correlation coefficient $R(OH, J(O^1D)) = 0.94$ including the instrument precision. One may presume indirect contributions to this result from OH reactants correlated with $J(O^1D)$. However, the correlation matrix (Supplementary Table 2) reveals that this is not the case. Furthermore, the monthly averages plotted in Fig. 2 demonstrate no detectable seasonal or annual trend of OH at Hohenpeissenberg apart from changes in $J(O^1D)$. We estimate the annual trend to be less than $\pm 2.5\%\ yr^{-1}$ (see Supplementary Fig. 2).

For a detailed variance analysis of the MOHp results (Fig. 3), the OH data were binned into series of time intervals of different lengths ('timescales') ranging from 5 minutes to 5 years. Four types of OH variances were calculated by averaging the variances of the respective time intervals: the total variance V_1 of observed OH, the variance V_2 common with $J(O^1D)$, the variance V_3 of instrument noise, and the variance $V_4 = V_1 - (V_2 + V_3)$ termed hereafter 'unassigned'. After subtracting the effect of instrument noise, we find that the variance of OH, relative to its final value at 5 years, is dominated by the diurnal cycle (76%) and the seasonal cycle (23%). The unassigned variance, V_4, containing information of the variability of chemical influences on OH, is zero at 5 minutes and reaches 10% at 10 days timescale, remaining constant above this range. This is the typical timescale of synoptic weather systems influencing the air mass composition in a specific region.

Such a strong relation between OH and $J(O^1D)$ can also be retrieved from the results of recent field campaigns of much shorter duration[17–20]. The campaigns were carried out at different locations, in quite diverse environments and with different instrumentation

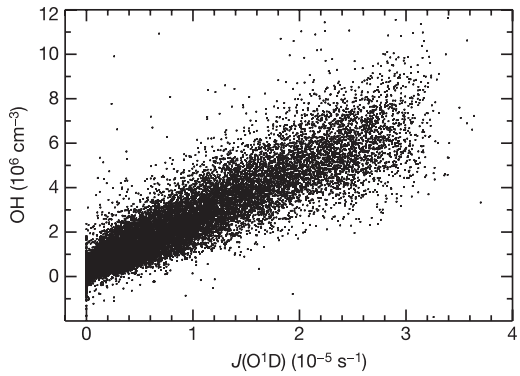


Figure 1 | Correlation of measured OH concentrations with simultaneously observed ozone photolysis frequencies, $J(O^1D)$. The data represent 5-min averages measured at the Meteorological Observatory Hohenpeissenberg between April 1999 and December 2003 ($N \approx 52{,}000$). A box-whisker representation of this data set is shown in Supplementary Fig. 1.

[1]Forschungszentrum Jülich, Institut ICG-II: Troposphäre, Jülich 52425, Germany. [2]German National Meteorological Service, DWD/MOHp, Hohenpeissenberg 82383, Germany.

© 2006 Nature Publishing Group

(Supplementary Table 3). Despite these large differences, the correlations between OH and $J(O^1D)$ were highly significant and compact (Supplementary Fig. 3). All data sets show the same functional dependence of OH on $J(O^1D)$ as the long-term MOHp data set, however, with different slopes, and in the case of the MINOS campaign[20], a different exponent. In general, this dependence can be described by an empirical power-law function:

$$[OH] = a \times (J(O^1D)/10^{-5}\,s^{-1})^b + c \quad (1)$$

The statistical analysis with respect to equation (1) for all of the campaigns is summarized in Supplementary Table 4. The results show that between 87% and 100% of the variance in observed OH is explained by the dependence of OH on $J(O^1D)$ and by instrument noise.

Because the production of OH via reactions $O_3 + h\nu \rightarrow O(^1D) + O_2$ and $O(^1D) + H_2O \rightarrow 2OH$ depends on the ozone photolysis frequency, it is plausible to expect a strong relation between $J(O^1D)$ and OH provided that other parameters such as O_3 and H_2O concentrations remain relatively constant. This was first demonstrated in ref. 21 and subsequently extended in ref. 22 to a broad range of NO_x levels. Empirically, the influence on OH of solar UV-B radiation is expressed in equation (1) using the term $J(O^1D)^b$. The exponent b reflects the combined effects of all photolytic processes—for example, the photolysis of O_3, NO_2, HONO, H_2O_2 and HCHO. Each of these processes generates OH either directly or via production of and recycling from HO_2, and all are highly correlated but not necessarily in a linear manner. The dependence of OH on reactants such as NO_x, hydrocarbons, O_3 or H_2O is condensed into the single pre-exponential coefficient, a. Finally, the coefficient c includes all processes that are light-independent—for example, OH production at night-time[23].

Coefficients a, b and c characterise the average influence of the chemical environment on OH at a specific location. The high variability in each of the photolytic processes and reactants (Supplementary Table 1) influencing OH may be expected to result in a strong variance around this average. However, in the present data we do not find such a strong variance. In contrast, the results show that although the individual reactants are highly variable, their combined influence on OH remains constant. Moreover, since this constancy is found even for relatively long time periods up to 5 years, the site specific values of the three coefficients a, b and c may well represent environmental conditions on larger spatial and temporal scales.

The results so far are entirely based on observations. The chemical environment of OH, the major cleansing agent of the troposphere, seems to be regulated in such a way that its relation to the driving force—solar radiation—is stabilized in a characteristic functional dependence. In a previous study[24] conducted in the lower stratosphere, a strong relation between OH and solar zenith angle was inferred. In principle, a similar analysis has been attempted here. However, very different chemical and meteorological conditions prevail in the troposphere. To better understand the OH–$J(O^1D)$ relation in the MOHp dataset, we performed a box-model calculation using a state-of-the-art chemistry module (see Methods) and the time series of long lived trace gases measured at MOHp[16]. The correlation between calculated OH levels and $J(O^1D)$ is characterised by $R = 0.912$ (Supplementary Fig. 4). Surprisingly, the results show that measured OH correlates more strongly with $J(O^1D)$ ($R = 0.941$, Fig. 1) than with calculated OH ($R = 0.925$, Supplementary Fig. 5). For the latter, an ideal model would yield $R = 0.98$ accounting only for experimental noise.

A detailed analysis of OH production and removal is shown in Fig. 4. The removal processes are subdivided into reactions with anthropogenic and biogenic trace gases, which exhibit quite different seasonal characteristics. However, the total of all removal processes is nearly constant. We also find a similar result for the OH production processes, which are scaled with $J(O^1D)$ in Fig. 4b. This normalization reveals the influence of parameters other than the photolysis frequencies. For example, ozone and water vapour exhibit maximum levels in summer, which is reflected in the scaled $(O^1D + H_2O)$ production term. The ratio between normalized production and removal processes shown in Fig. 4c is equivalent to the slope a for the OH–$J(O^1D)$ relation in equation (1). As neither the sum of the production terms or removal terms of OH in the model exhibit a significant seasonality, their ratio is also quasi-constant. The corresponding slope derived from the OH and $J(O^1D)$ measurements shows the same seasonal stability, but is on average 30% smaller. This deviation is well covered by the accuracy of the chemical ionization mass spectrometer (CIMS) used to measure OH concentration (20%), and the estimated error of the model calculation (21%, see Supplementary Tables 5 and 6).

By simplifying the reaction scheme for the chemical regime at Hohenpeissenberg (see Methods, equations (2)–(18)) we further show how the OH–$J(O^1D)$ relation results in a linear dependence. This approach is complemented by a sensitivity analysis (Supplementary Table 5), which quantifies how calculated OH is influenced by the boundary conditions (levels of NO_x, CO, O_3 and so on). Except for the combined photolysis frequencies, all parameters have very much attenuated influence on OH. For example, NO_x has a sensitivity coefficient of 0.2, which means that scaling of NO_x with a factor 2 is transformed into an OH change by a factor of $2^{0.2} = 1.15$.

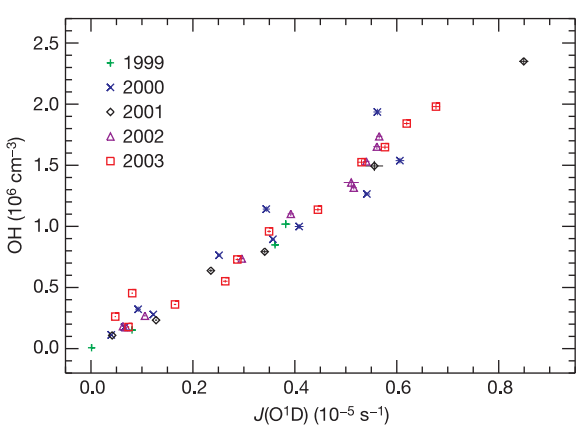


Figure 2 | Monthly averages of simultaneous observations of OH and $J(O^1D)$ measured in different years at MOHp. Standard deviations of the means are in the size range of the symbols. The correlation coefficient between monthly averages of OH and $J(O^1D)$ is 0.985, $N = 41$.

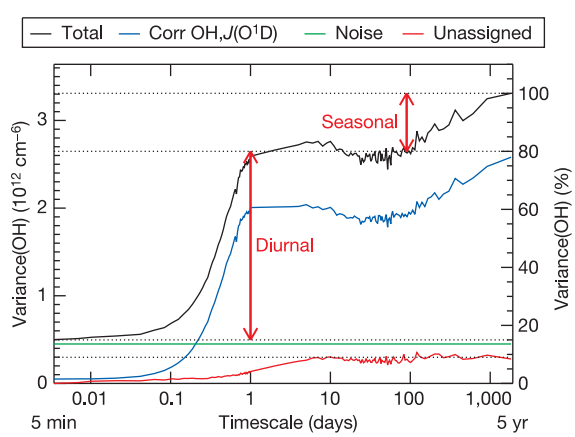


Figure 3 | Variance analysis of the 30-s time-resolved OH data measured at MOHp for 1999–2003. The black line denotes the total variance of observed OH (V_1), the blue line the partial variance of OH common with $J(O^1D)$ (V_2), the green line the variance of instrument precision (V_3), and the red line the unassigned variance of OH (V_4) (see main text and Methods).

© 2006 Nature Publishing Group

On the basis of our preceding analysis, both model calculations and measurements show a high degree of seasonal stability in the relation between OH and $J(O^1D)$. However, the simple empirical relation established in equation (1) correlates with OH measurements better than the detailed photochemistry model, which uses some 50 additional measured parameters. Therefore, $J(O^1D)$ multiplied by a slope characterizing the corresponding chemical regime appears to describe the OH concentration better than the detailed model, at least for the current dataset (see Methods).

We conclude that scaling of OH by $J(O^1D)$, which depends on solar UV radiation, eliminates most of the diurnal and seasonal variation, and transforms OH into a parameter with significantly reduced variability. We propose that regional or even global OH distributions can be characterized by a simple set of coefficients for timescales on the order of months or even years. This approach may be used to define an 'OH index' that describes observable impacts and trends in the oxidation efficiency of the troposphere in different chemical regimes. For the chemical regime represented by the Hohenpeissenberg data, our corresponding analysis has shown no observable trend in OH levels between 1999 and 2003.

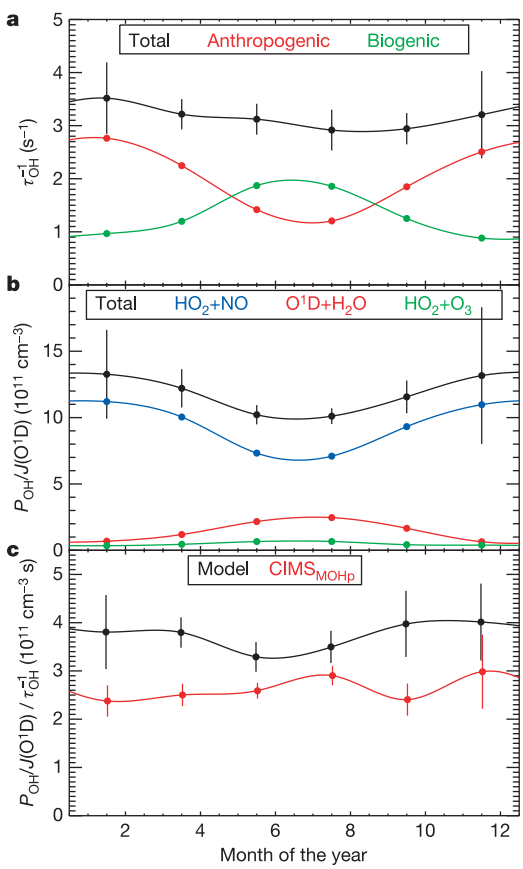


Figure 4 | Bimonthly averages of destruction and production processes of OH at MOHp. Error bars denote the respective standard deviations. Production and destruction processes (**a** and **b**) are calculated from co-located measurements of OH precursors and reactants. Destruction is described by anthropogenic (reaction with CO, NO_2, NO, HNO_3, ethene) and biogenic (all other reactions, mainly with CH_4, HCHO, isoprene) reaction frequencies, τ_{OH}^{-1}. Production rates, P_{OH}, are scaled with $J(O^1D)$. The ratio of production and destruction processes scaled with $J(O^1D)$ in steady-state (PSS) is equal to the slope *a* of [OH] versus $J(O^1D)$ ($a = [OH]/J(O^1D)$; $[OH]_{PSS}$ = OH production rate/inverse lifetime of OH = P_{OH}/τ_{OH}^{-1}). Panel **c** compares bi-monthly means of model calculations of $(P_{OH}/J(O^1D))/\tau_{OH}^{-1}$ and of slope *a* calculated from OH measurements made using the chemical ionization mass spectrometer, $CIMS_{MOHp}$.

METHODS

Experimental. OH was measured online with 30-s time resolution by derivatization with sulphur dioxide and detection of the corresponding sulphuric acid product using chemical ionization mass spectrometry[25]. For the data set considered here (April 1999–December 2003), we have analysed the differences between each 30-s data point and its most adjacent data points and estimate an average OH detection limit of 1.4×10^5 cm^{-3} (2 standard deviations), a measurement precision (1 standard deviation) of 0.7×10^5 cm^{-3} + 0.13[OH] based on 5-min signal integration, and a measurement accuracy (1 standard deviation) of 20%. The photolysis frequency, $J(O^1D)$, was measured with two filter-radiometers (2π sr each[26]) at 1-min time resolution with an accuracy of 15% (for zenith angles <75°) and a <0.1% precision.

Model description. For the numerical description of OH, the corresponding reaction scheme RACM[27] is used with advanced isoprene chemistry[28] in a box model calculation. Concentrations of the long lived OH reactants O_3, NO, NO_2, CO, isoprene, C_2–C_7 alkanes and alkenes, acetylene, propyne, benzene, toluene, ethylbenzene and xylenes were concurrently measured at the GAW station Hohenpeissenberg[16] together with ambient temperature, pressure, dew point, $J(NO_2)$ and $J(O^1D)$. Methane and H_2 mixing ratios are assumed to be 1.9 p.p.m. (parts per million) and 550 p.p.b. (parts per billion), respectively. The calculation is performed in a steady-state mode with an additional lifetime of 24 h for the reaction products of the measured species to avoid the build-up of unrealistic concentrations of secondary products.

Generalized reaction scheme. A reaction system describing the generalized structure of OH photochemistry can be written as follows: (rate coefficients[27] given in parenthesis refer to 298 K, 1,013 hPa, 10 hPa H_2O):

$$O_3 + h\nu(+H_2O) \rightarrow 2OH \quad (\alpha J(O^1D)) \tag{2}$$

$$OH + \text{hydrocarbons, CO, } H_2 \rightarrow HO_2 + \text{products} \ (\tau_{HC}^{-1} = 3.3\text{s}^{-1}) \tag{3}$$

$$HO_2 + NO \rightarrow OH + NO_2 \ (k_{HO_2+NO} = 8.56 \times 10^{-12}\ \text{cm}^3\ \text{s}^{-1}) \tag{4}$$

$$OH + NO_2 \rightarrow HNO_3 \ (k_{OH+NO_2} = 11.5 \times 10^{-12}\ \text{cm}^3\ \text{s}^{-1}) \tag{5}$$

$$HO_2 + HO_2(+H_2O) \rightarrow H_2O_2 \ (k_{HO_2+HO_2} = 4.5 \times 10^{-12}\ \text{cm}^3\ \text{s}^{-1}) \tag{6}$$

$$NO + O_3 \rightarrow NO_2 \ (k_{NO+O_3} = 1.82 \times 10^{-14}\ \text{cm}^3\ \text{s}^{-1}) \tag{7}$$

$$NO_2 + h\nu \rightarrow NO + O_3 \quad (J(NO_2)) \tag{8}$$

here

$$\alpha = (k_{O^1D+H_2O}[H_2O])/(k_{O^1D+H_2O}[H_2O] + k_{O^1D+N_2}[N_2] + k_{O^1D+O_2}[O_2]) \tag{9}$$

The balance equations for the steady-state levels of OH, HO_2, and NO are:

$$2\alpha J(O^1D)[O_3] + [HO_2][NO]k_{HO_2+NO} = [OH]\tau_{HC}^{-1} + [OH][NO_2]k_{OH+NO_2} \tag{10}$$

$$[OH]\tau_{HC}^{-1} = [HO_2][NO]k_{HO_2+NO} + 2[HO_2]^2 k_{HO_2+HO_2} \tag{11}$$

$$[NO][O_3]k_{NO+O_3} + [HO_2][NO]k_{HO_2+NO} = J(NO_2)[NO_2] \tag{12}$$

which can be combined to:

$$[NO] = (J(NO_2)[NO_2])/([O_3]k_{NO+O_3} + [HO_2]k_{HO_2+NO}) \tag{13}$$

$$2[HO_2]^2 k_{HO_2+HO_2}\tau_{HC} + [HO_2][NO]k_{HO_2}\tau_{HC} = (2\alpha J(O^1D)[O_3] + [HO_2][NO]k_{HO_2+NO})/(\tau_{HC}^{-1} + [NO_2]k_{OH+NO_2}) \tag{14}$$

For average conditions at Hohenpeissenberg, $[NO_2]k_{OH+NO_2}$ is small compared to τ_{HC}^{-1} and $[HO_2]k_{HO_2+NO}$ is small compared to $[O_3]k_{NO+O_6}$. Neglecting both terms yields the following expressions, similar to the derivation in ref. 29:

$$[HO_2] = \sqrt{\frac{\alpha J(O^1D)[O_3]}{k_{HO_2+HO_2}}} \tag{15}$$

$$[OH] = \sqrt{\frac{\alpha J(O^1D)[O_3]}{k_{HO_2+HO_2}}} \times \frac{J(NO_2)[NO_2]}{[O_3]k_{NO+O_3}} \times k_{HO_2+NO}\tau_{HC} + 2\alpha J(O^1D)[O_3]\tau_{HC} \tag{16}$$

The product $2\alpha J(O^1D)[O_3]$ in equation (16) is small at Hohenpeissenberg

© 2006 Nature Publishing Group

compared to the recycling by way of $HO_2 + NO$ (see Fig. 4) and (as in other observations[22]) $J(NO_2)$ is observed to correlate with $\sqrt{J(O^1D)}$:

$$J(NO_2) = F_J\sqrt{J(O^1D)} \quad (17)$$

This results in the final expression for OH with respect to Hohenpeissenberg:

$$[OH] = \frac{k_{HO_2+NO}\tau_{HC}[NO_2]F_J}{k_{NO+O_3}} \times \sqrt{\frac{\alpha}{k_{HO_2+HO_2}[O_3]}} \times J(O^1D) \quad (18)$$

For Hohenpeissenberg, $\alpha = 0.075$ and $F_J = 2\,s^{-0.5}$. Note that the influence of both O_3 and H_2O on OH is reduced compared to the primary production rate $2\alpha J(O^1D)[O_3]$: in the case of O_3, to $(1/\sqrt{O_3})$, and in case of H_2O to $\sqrt{\alpha/k_{HO_2+HO_2}}$.

In summary, the observed strong relation to $J(O^1D)$ is a direct consequence of the efficient recycling of OH by way of the reaction $HO_2 + NO$. The hydroperoxy radical, HO_2, shows a square root dependence on $J(O^1D)$. As NO is formed by photolysis of NO_2 (equation (8)) with rate coefficient $J(NO_2)$, NO is proportional to $J(NO_2)$, which strongly correlates with the square root of $J(O^1D)$. Therefore both HO_2 and NO show square root dependencies on $J(O^1D)$ which combine to an overall linear dependence of OH on $J(O^1D)$, as shown in equation (18).

Precision of measured and calculated OH concentrations. At 5-min time resolution, the CIMS instrument has an estimated precision of $e_{CIMS} = 4\%$ (variance of the error relative to the total variance of OH). It should have a correlation coefficient $R = (1 - 0.04)^{0.5} = 0.979$ with the true OH values if the true values have no correlation with the noise of the CIMS instrument. A correlation coefficient of 0.941 between OH_{CIMS} and $J(O^1D)$ implies an error of $e_{J(O^1D)} = 7.8\%$ relative to the total variance of OH for the calculation of OH using $J(O^1D)$. It is calculated from $R^2 = (1 - e_{CIMS})(1 - e_{JO^1D})$. This is correct if the CIMS and $J(O^1D)$ measurements have uncorrelated errors. As the variance of OH adjusted to the instrument noise of the CIMS instrument is near zero at 5-min timescale, and the error of the $J(O^1D)$ calculation is generated at timescales of 10 days (unassigned variance in Fig. 3), the precision of the CIMS instrument reduces to 8% at a timescale of 2.5 min but the precision of the $J(O^1D)$-calculation of OH remains 7.8%. With the same formalism, the precision of the RACM model for the prediction of OH is $e_{model} = 11\%$ (calculated from $R^2 = (1 - e_{CIMS})(1 - e_{model}) = 0.925^2$).

Variance analysis. For the variance analysis shown in Fig. 3, the 30-s time-resolved OH data were used and divided recursively into time intervals of different lengths. For example, with respect to a timescale of 1 day, the 5-year data set was divided into 1,826 consecutive intervals. For each of these timescales, four types of OH variances were calculated by averaging the variances of the respective time intervals (var = variance, obs = observed): (1) the total variance of observed OH, $V_1 = \mathrm{var}(OH_{obs})$; (2) the variance common with $J(O^1D)$, $V_2 = \mathrm{var}(\mathrm{Cor}\ OH, J(O^1D)) = R^2(OH_{obs}, J(O^1D)^b)\mathrm{var}(OH_{obs})$; (3) the variance of instrument noise, $V_3 = \mathrm{var}(\text{instrument noise})$; and (4) and the unassigned variance, $V_4 = V_1 - (V_2 + V_3)$.

The terms 'Instrument precision' and 'Explained variance in OH' in Supplementary Table 4 denote V_3/V_1 and $(V_2 + V_3)/V_1$, respectively, calculated for a time scale of 5 years. 'Explained variance in OH' is used from a statistical point of view. It indirectly includes the effect of all other influence factors for OH that are correlated with $J(O^1D)$.

Received 13 December 2005; accepted 19 May 2006.

1. Levy, H. II Normal atmosphere: Large radical and formaldehyde concentrations predicted. *Science* **173**, 141–143 (1971).
2. Ehhalt, D. H. Radical ideas. *Science* **279**, 1002–1003 doi:10.1126/science.279.5353.1002 (1998).
3. Jacob, D. J. in *Handbook of Weather, Climate and Water* (eds Potter, T. D. & Colman, B. R.) 29–46 (Wiley & Sons, New York, 2003).
4. Lelieveld, J., Dentener, F. J., Peters, W. & Krol, M. C. On the role of hydroxyl radicals in the self-cleansing capacity of the troposphere. *Atmos. Chem. Phys.* **4**, 2337–2344 (2004).
5. Prinn, R. G. *et al.* Evidence for substantial variations of atmospheric hydroxyl radicals in the past two decades. *Science* **292**, 1882–1888 (2001).
6. Manning, M. R., Lowe, D. C., Moss, R. C., Bodeker, G. E. & Allan, W. Short-term variations in the oxidizing power of the atmosphere. *Nature* **436**, 1001–1004 (2005).
7. Perner, D. *et al.* Measurements of tropospheric OH concentrations: A comparison of field data with model predictions. *J. Atmos. Chem.* **5**, 185–216 (1987).
8. Platt, U., Rateike, M., Junkermann, W., Rudolph, J. & Ehhalt, D. H. New tropospheric OH measurements. *J. Geophys. Res.* **93**, 5159–5166 (1988).
9. Dorn, H.-P., Callies, J., Platt, U. & Ehhalt, D. H. Measurement of tropospheric OH concentrations by laser long-path absorption spectroscopy. *Tellus B* **40**, 437–445 (1988).
10. Eisele, F. L. & Tanner, D. J. Ion-assisted tropospheric OH measurements. *J. Geophys. Res.* **96**, 9295–9308 (1991).
11. Wennberg, P. O. *et al.* In situ measurements of OH and HO_2 in the upper troposphere and stratosphere. *J. Atmos. Sci.* **52**, 3413–3420 (1995).
12. Crosley, D. R. The measurement of OH and HO_2 in the atmosphere. *J. Atmos. Sci.* **52**, 3299–3314 (1995).
13. Brune, W. H., Stevens, P. S. & Mather, J. H. Measuring OH and HO_2 in the troposphere by laser-induced fluorescence at low pressure. *J. Atmos. Sci.* **52**, 3328–3336 (1995).
14. Heard, D. E. & Pilling, M. J. Measurement of OH and HO_2 in the troposphere. *Chem. Rev.* **103**, 5163–5198 (2003).
15. World Meteorological Organisation *Strategy for the Implementation of the Global Atmosphere Watch Programme (2001–2007)* (GAW Rep. No. 142, WMO TD No 1077, WMO, Geneva, 2001)
16. Deutscher Wetterdienst, Global Atmosphere Watch. ⟨http://www.dwd.de/gaw⟩ (2002).
17. Holland, F., Aschmutat, U., Heßling, M., Hofzumahaus, A. & Ehhalt, D. H. Highly time resolved measurements of OH during POPCORN using laser-induced fluorescence spectroscopy. *J. Atmos. Chem.* **31**, 205–225 (1998).
18. Brauers, Th., Hausmann, M., Bister, A., Kraus, A. & Dorn, H.-P. OH radicals in the boundary layer of the Atlantic Ocean 1. Measurements by long-path laser absorption spectroscopy. *J. Geophys. Res.* **106**, 7399–7414 (2001).
19. Holland, F., Hofzumahaus, A., Schäfer, J., Kraus, A. & Pätz, H.-W. Measurements of OH and HO_2 radical concentrations and photolysis frequencies during BERLIOZ. *J. Geophys. Res.* **108**, 8246, doi:10.1029/2001JD001393 (2003).
20. Berresheim, H., Plass-Dülmer, C., Elste, T., Mihalopoulos, N. & Rohrer, F. OH in the coastal boundary layer of Crete during MINOS: Measurements and relationship with ozone photolysis. *Atmos. Chem. Phys.* **3**, 639–649 (2003).
21. Ehhalt, D. H., Dorn, H.-P. & Poppe, D. The chemistry of the hydroxyl radical in the troposphere. *Proc. R. Soc. Edinb. B* **97**, 17–34 (1991).
22. Ehhalt, D. H. & Rohrer, F. Dependence of the OH concentration on solar UV. *J. Geophys. Res.* **105**, 3565–3571 (2000).
23. Faloona, I. *et al.* Nighttime observations of anomalously high levels of hydroxyl radicals above a deciduous forest canopy. *J. Geophys. Res.* **106** (D20), 24315–24333 (2001).
24. Hanisco, T. F. *et al.* Sources, sinks, and the distribution of OH in the lower stratosphere. *J. Phys. Chem.* **105**, 1543–1553 (2001).
25. Berresheim, H., Elste, T., Plass-Dülmer, C., Eisele, F. L. & Tanner, D. J. Chemical ionization mass spectrometer for long-term measurements of atmospheric OH and H_2SO_4. *Int. J. Mass Spectrom.* **202**, 91–109 (2000).
26. Bohn, B., Kraus, A., Müller, M. & Hofzumahaus, A. Measurement of atmospheric $O_3 \rightarrow O(^1D)$ photolysis frequencies using filterradiometry. *J. Geophys. Res.* **109**, D10S90, doi: 10.1029/2003JD004319 (2004).
27. Stockwell, W. R., Kirchner, F., Kuhn, M. & Seefeld, S. A new mechanism for regional atmospheric chemistry modelling. *J. Geophys. Res.* **102**, 25847–25879 (1997).
28. Geiger, H., Barnes, I., Bejan, J., Benter, T. & Spittler, M. The tropospheric degradation of isoprene: an updated module for the regional atmospheric chemistry mechanism. *Atmos. Environ.* **37**, 1503–1519 (2003).
29. Sillman, S., Logan, J. A. & Wofsy, S. The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes. *J. Geophys. Res.* **95**, 1837–1851 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Eisele and D. Tanner for help in setting up the CIMS system at MOHp, T. Elste and G. Stange for OH measurements, and E. Tensing for J (O^1D) measurements, D. H. Ehhalt, A. Wahner and C. Plass-Dülmer for discussions, the GAW team at MOHp for the ancillary data, and DWD/BMVBS for financial support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Requests for the Hohenpeissenberg data sets should be addressed to H.B. (harald.berresheim@dwd.de). Correspondence and requests for materials should be addressed to F.R. (f.rohrer@fz-juelich.de).

© 2006 Nature Publishing Group

LETTERS

Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip

David R. Shelly[1], Gregory C. Beroza[1], Satoshi Ide[2] & Sho Nakamula[3]

Non-volcanic seismic tremor was discovered in the Nankai trough subduction zone in southwest Japan[1] and subsequently identified in the Cascadia subduction zone[2]. In both locations, tremor is observed to coincide temporally with large, slow slip events on the plate interface downdip of the seismogenic zone[2–7]. The relationship between tremor and aseismic slip remains uncertain, however, largely owing to difficulty in constraining the source depth of tremor. In southwest Japan, a high quality borehole seismic network allows identification of coherent S-wave (and sometimes P-wave) arrivals within the tremor, whose sources are classified as low-frequency earthquakes. As low-frequency earthquakes comprise at least a portion of tremor, understanding their mechanism is critical to understanding tremor as a whole. Here, we provide strong evidence that these earthquakes occur on the plate interface, coincident with the inferred zone of slow slip. The locations and characteristics of these events suggest that they are generated by shear slip during otherwise aseismic transients, rather than by fluid flow. High pore-fluid pressure in the immediate vicinity, as implied by our estimates of seismic P- and S-wave speeds, may act to promote this transient mode of failure. Low-frequency earthquakes could potentially contribute to seismic hazard forecasting by providing a new means to monitor slow slip at depth.

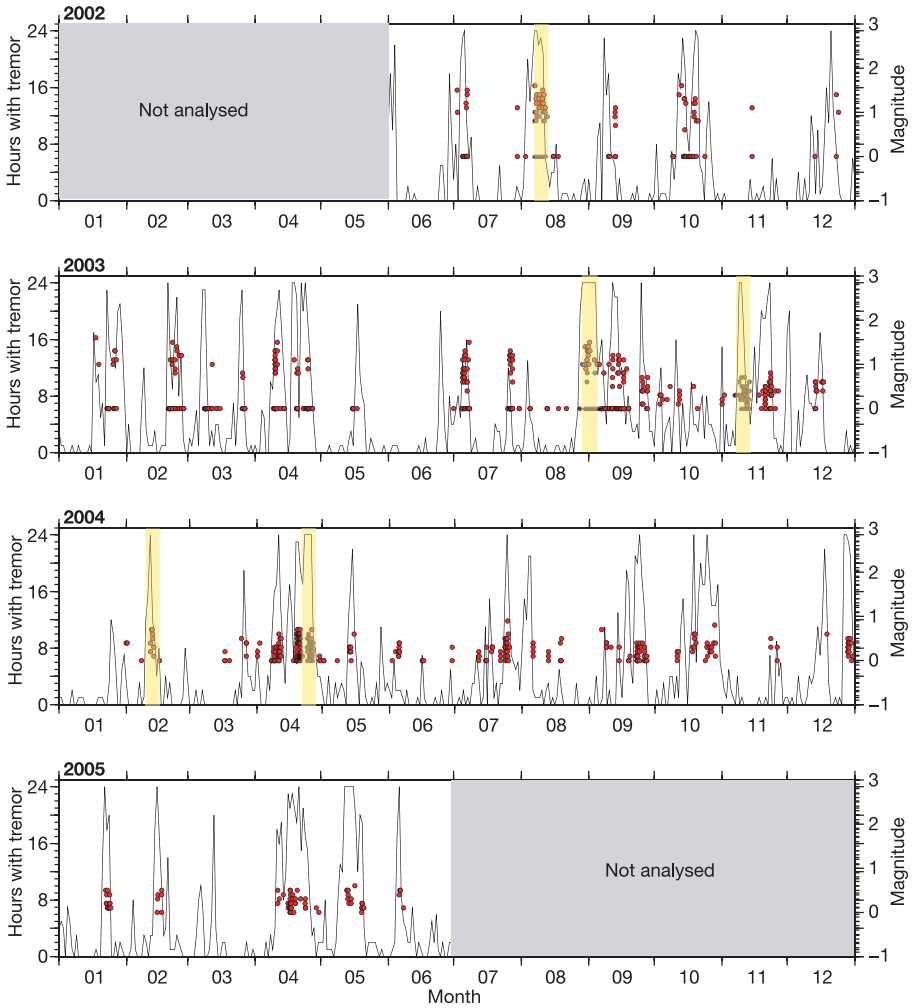


Figure 1 | Comparison of temporal distribution of tremor activity and low-frequency earthquakes (LFEs) in western Shikoku. The horizontal axis indicates time, with each year a separate panel, beginning in 2002. Red dots denote LFEs with magnitude on the vertical axis. The black line indicates the number of hours containing tremor per day. Yellow bars indicate the period and duration of detected slow slip events[5,6], which are modelled to occur near the plate boundary beneath western Shikoku. Events with an undetermined magnitude are plotted as magnitude zero. The magnitude determination method changes during September 2003.

[1]Department of Geophysics, 397 Panama Mall, Stanford University, Stanford, California 94305-2215, USA. [2]Department of Earth and Planetary Science, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan. [3]Earthquake Research Institute, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-0032, Japan.

A key to understanding the mechanism of deep non-volcanic tremor is determining exactly where it is generated; however, tremor is very difficult to locate owing to its small amplitude and the scarcity of coherent impulsive wave arrivals of the sort used to locate ordinary earthquakes. Epicentral locations based on seismogram envelope cross-correlation constrain tremor to lie within a belt that follows the strike of the subducting plate[2]. At these locations the depth of the plate interface is 30–45 km, but the depth of the tremor source is less certain.

At least two hypotheses might explain the observed tremor signals. One is that tremor is generated by the movement of fluids at depth, either by hydraulic fracturing[1,8] or by coupling between the rock and fluid flow[9]. In this model, tremor might be analogous to much shallower volcanic tremor, which is thought to arise from the coupling of fluid flow to the solid Earth in volcanic systems[10,11]. The correlation between tremor and aseismic slip could be explained if the slip is triggered by the same fluid movement that generates the tremor or, alternatively, if the fluid flow is a response to changes in stress and strain induced by slip[12].

A second possibility is that tremor is generated directly by slow shear slip on the plate interface[7]. Under this hypothesis, tremor is the weak seismological signature of slip that is otherwise too slow to generate detectable seismic waves. In this scenario, tremor could be generated during aseismic transients as slip locally accelerates owing to the effects of geometric or physical irregularities on the plate interface. Fluids might play an auxiliary role, altering the conditions on the plate interface to enable transient slip events, without generating seismic waves directly.

Since 1999, the Japan Meteorological Agency (JMA) has differentiated a new class of events denoted as low-frequency earthquakes (LFEs) in their seismicity catalogue. These events occur almost exclusively as part of an extended duration tremor signal[7,9] (Fig. 1). They therefore correlate strongly with observed slow slip events. The LFEs and extended tremor also locate in approximately the same region[1,6–9] and exhibit similar migration behaviour along-strike[1,5–7]. Although it is uncertain whether or not LFEs and extended duration tremor represent a single phenomenon, their close association means that their mechanisms are probably intertwined. Figure 2 shows waveform records of a long-duration tremor sequence containing several LFEs. The visible arrivals from these events are primarily S waves, and if arrivals from the same source are distinguished at enough stations, the event is included in the JMA seismic catalogue. For such events, usually only the S-wave arrival time is reported. Owing to the lack of P-wave measurements and because the S-wave arrival is often emergent, catalogue locations of these events have large uncertainties, especially in depth.

To improve this situation, we use a combination of waveform cross-correlation and double-difference tomography[13,14] to relocate these LFEs in the western Shikoku region of the Nankai trough. We also obtain high-resolution velocity structure and precise locations of ordinary earthquakes. Although tremor (including LFEs) is observed along much of the Nankai trough, it is exceptionally active in western Shikoku.

Figure 3 shows the events, stations and the horizontal distribution of velocity inversion nodes used in this study. Vertically, the velocity node spacing is 5 km. In total, we relocate 6,713 events occurring between June 2002 and June 2005, including 1,180 LFEs, using waveforms from 112 seismic stations operated by the National Research Institute for Earth Science and Disaster Prevention (NIED), JMA, University of Tokyo and Kochi University. This includes 77 stations from the recently installed NIED Hi-net high-sensitivity borehole network, which provides especially high quality, low noise waveforms[15]. In addition to these data, we use P- and S-wave arrival time picks, provided by JMA for all stations. More details about the data and inversion process are included in Supplementary Information.

We find that the waveforms of LFEs recorded at the same station often correlate with each other, suggesting a similar source location and mechanism for these events. Their waveform similarity is high enough for successful relocation using cross-correlation-derived differential times. Starting with only a rough estimate of the P- or S-wave arrival time, we obtain precise differential travel time measurements by cross-correlation waveform alignment. Supplementary Fig. S1 demonstrates the dramatic reduction in measurement error achieved by cross-correlation of LFE waveforms. Among the LFEs, S-wave correlations are particularly successful owing to the greater amplitude and signal to noise ratio of the S waves compared with P waves. However, despite the lack of P-wave arrival times in the catalogue, we are able to obtain P-wave differential arrival times for LFEs by correlating waveforms in a window centred on a theoretical P arrival time estimated from the S-wave arrival time and origin time (Supplementary Fig. S2). We are also able to obtain successful P-wave and S-wave correlations between some LFEs and ordinary earthquakes

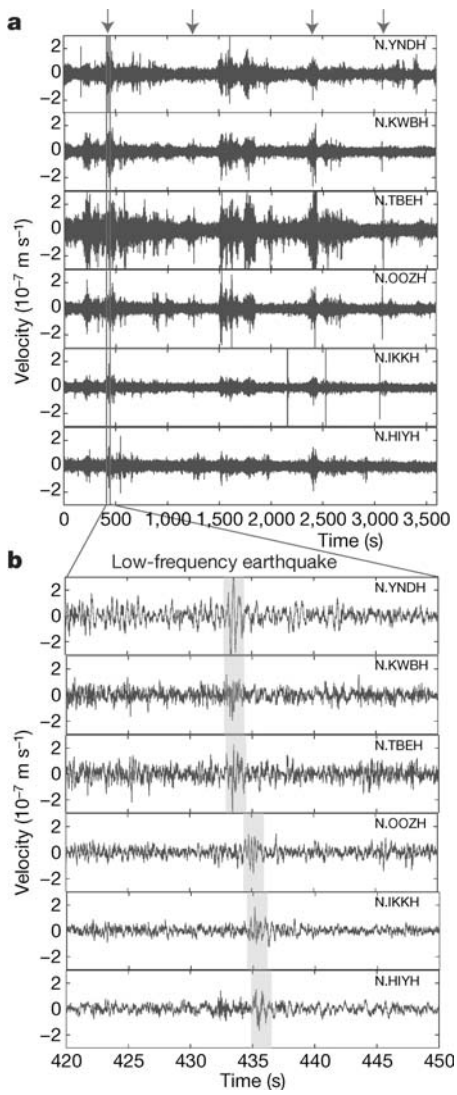


Figure 2 | An extended period of tremor containing four identified LFEs. **a**, One hour of unfiltered east component velocity seismograms for 29 August 2005, beginning at 17:00 local time during a tremor sequence as recorded at six Hi-net borehole stations (N.YNDH and so on) in western Shikoku. Arrows indicate times of LFEs. **b**, Expanded view of seismograms showing one LFE from **a**. S-wave arrivals can be identified on these seismograms between 433 s and 436 s and are highlighted in grey. Traces are ordered by the time of the S-wave arrival, with the earliest arrival on top. P-wave arrival is too weak to be obviously visible above the noise, even on vertical seismograms, but often can be measured for such events using cross-correlation.

(Supplementary Fig. S3). These correlations are especially important because they demonstrate similarity of the LFEs with known shear sources, and because they help locate the LFEs relative to intra-slab seismicity. The correlations with regular seismicity along with the observed larger S-wave amplitudes suggest that the LFEs themselves represent shear failure.

After relocation, the LFEs collapse onto a distinct plane 5–8 km above, and approximately parallel to, the dipping plane of seismicity within the subducting Philippine Sea slab (Fig. 4). In places, the plane of LFEs extends for nearly 20 km in the dip direction of the slab, with events strongly clustered along-strike (Fig. 3). Although it was previously thought that these events occurred within the overriding plate[9], we believe they occur on the plate interface. This hypothesis is supported by an independent seismic refraction study, which finds that the plate interface is 3–5 km shallower than previous estimates[16], with the top surface corresponding closely to our LFE locations.

Assuming average oceanic crustal thickness (8 km), the location of LFEs on the plate interface places the intra-slab seismicity primarily within the oceanic lower crust (Fig. 4d). An increase in P- and S-wave velocities seen just below the slab seismicity (Fig. 4), as well as recent receiver function analysis[17], further support this location of the oceanic Moho. The intra-slab seismicity appears to begin within the lowermost crust at shallower depths, and to expand upward within the crust as the slab subducts. Such behaviour is consistent with thermal-petrologic models for this region, which predict metamorphic reactions to occur first within the hotter lower crust[18], but contrasts with expectations that intra-slab seismicity occurs primarily within the upper crust[19]. This issue remains an important subject of future research.

In the immediate vicinity of the LFEs, we find a zone of high v_P/v_S ratio (1.9–1.95, where v_P and v_S are respectively P- and S-wave velocity), probably indicating high pore-fluid pressures[20]. Figure 4 shows the estimated P, S and v_P/v_S velocity structure on a cross-section through an area of high LFE activity. A chequerboard test (Supplementary Fig. S4) demonstrates that our model is well resolved. The most prominent area of high v_P/v_S is located between the LFEs and the intra-slab seismicity, within the subducting crust. The high-v_P/v_S zone is most pronounced on the cross-section with the most LFE activity (Fig. 4), and is consistent with the hypothesis that these LFEs are enabled by the release of fluids in the subduction zone. A likely source of fluids is dehydration of hydrous minerals within the subducting oceanic crust, a process also proposed to enable intra-slab seismicity[19,21,22].

High pore-fluid pressure on the plate interface has the potential to alter the mode of slip by extending the conditionally stable region, as proposed for the Tokai segment at the northeast end of the Nankai trough[23] in order to explain a long-term slow slip event on this segment[24]. Interestingly, this region also experiences a high level of tremor activity[1], with epicentral locations of the tremor matching a region of high v_P/v_S within the subducting crust[23]. This

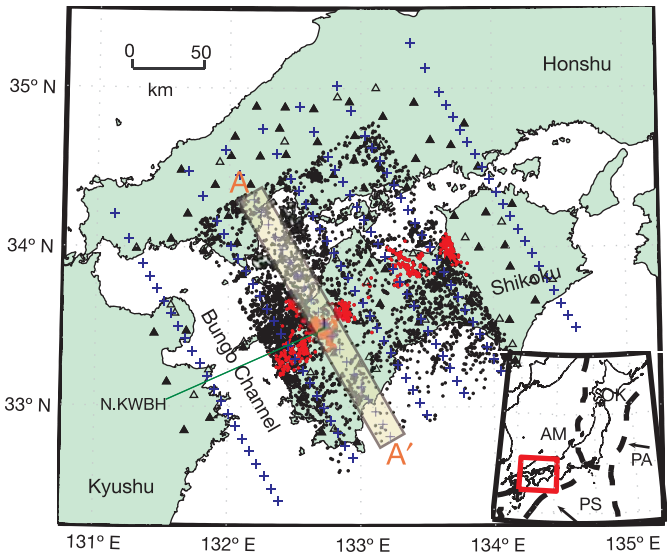


Figure 3 | Area map with events, stations and velocity nodes used in the inversion. Black dots indicate regular earthquakes and red dots indicate LFEs. Open black triangles show seismic stations used in the inversion, filled black triangles denote Hi-net borehole stations, and blue plus signs indicate the horizontal locations of velocity nodes. The shaded yellow region shows the zone of earthquakes and LFEs plotted on cross-section A–A′ in Fig. 4. Inset at lower right shows the broader tectonics of the region, with the red box indicating the area shown in the main figure. Dashed lines indicate plate boundaries. PA, Pacific plate; PS, Philippine Sea plate; AM, Amur plate; OK, Okhotsk plate. Plate model is PB2002[30]. Also shown is the location of station N.KWBH referred to in Supplementary Figs S1–S3.

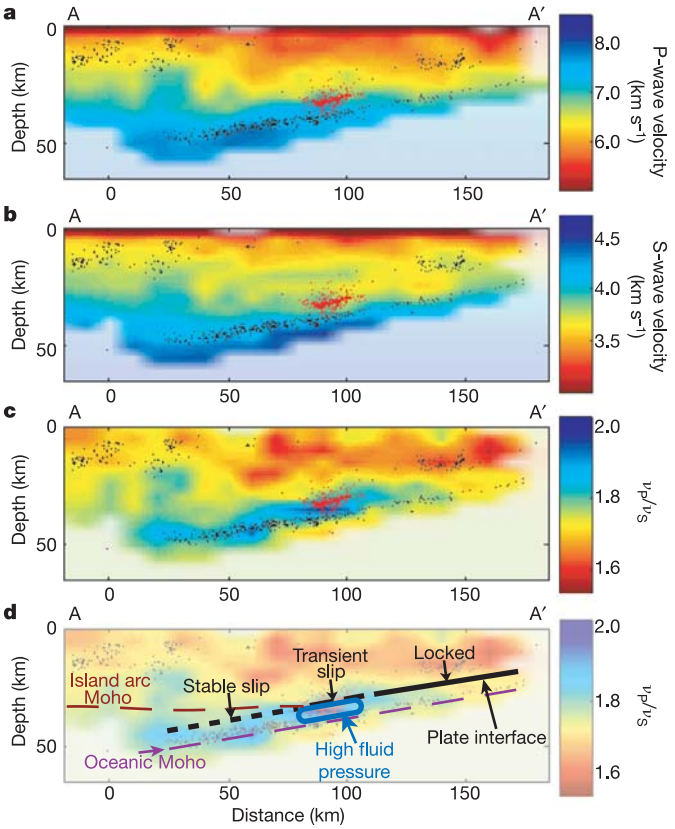


Figure 4 | Seismic velocities and relocated hypocentres along cross-section A–A′. **a**, P-wave velocity, v_P. **b**, S-wave velocity, v_S. **c**, v_P/v_S ratio. This cross-section is selected because it cuts through the region of highest tremor and LFE activity (see Fig. 2). Relocated hypocentres are plotted for events within 9 km of the cross-section (the yellow highlighted region in Fig. 2), with black dots indicating regular earthquakes and the red dots indicating LFEs. Faded areas represent regions of poor resolution, where the derivative weight sum (DWS) value is less than 50. **d**, Schematic diagram of our interpretation overlayed on the v_P/v_S structure. We interpret the LFEs to occur on the plate interface, coincident with the zone of transient slip, while the seismicity within the slab occurs primarily within the oceanic lower crust. The oceanic Moho is drawn based on an 8 km crustal thickness, which is consistent with observed P- and S-wave velocities in **a** and **b**. Between the LFEs and slab seismicity, within the slab crust, is a distinct region of high v_P/v_S ratio (1.9–1.95), which we interpret as due to high fluid pressure. The plate interface updip of the LFEs is apparently locked, based on rupture models of the 1946 Nankai earthquake[16,26] and the lack of current seismicity. Downdip of the LFEs, the plate interface may slip stably. Also shown is the approximate location of the island arc Moho.

correspondence between active tremor and high v_P/v_S is strikingly similar to what we observe in western Shikoku.

Tremor has recently been discovered outside a subduction environment under the San Andreas fault system south of Parkfield[25]. This tremor, estimated to occur below the seismogenic zone at depths of 20–40 km, shares many characteristics with that observed in Japan and Cascadia. Like tremor in subduction zones, tremor on the San Andreas fault occurs in a region previously thought to deform aseismically, but no associated deformation transient has yet been detected.

We propose that the coupled phenomena of tremor, LFEs and episodic slow slip events represent a mode of failure for a transition zone between a locked and continuously creeping fault. In a subduction zone, this corresponds to the transition between the locked megathrust source region updip and the continuously slipping region downdip. In fact, models of co-seismic rupture for the 1946 M_w (moment magnitude) 8.2 Nankai earthquake[16,26] terminate only a short distance updip of the observed band of LFEs. Some areas where tremor and LFEs occur, such as the Tokai (discussed above) and Bungo Channel regions, have also been observed to fail in longer duration slow slip events[24,27,28], demonstrating a tendency for transient slip in these areas.

Precise locations indicate that the LFEs analysed in this study occur on the plate interface. The correlation between LFEs and regular seismicity, the fact that LFEs are primarily composed of shear waves, and the remarkable correspondence with slow slip events[5–7] argue that LFEs represent shear slip on this interface. We propose that LFEs may be generated by local slip accelerations at geometric or frictional heterogeneities that accompany large slow slip events on the plate interface. Long-duration tremor may result simply from a superposition of many concurrent LFEs. Alternatively, long-duration tremor might represent a combined signal of shear slip and fluid flow. We hypothesize that increasing fluid pressure may reduce the effective normal stress and enable slip on the plate interface; this slip could increase permeability and allow pressurized fluid to escape, possibly contributing to the tremor signal in the process. In either case, evidence discussed above suggests that fluids play a key role in the failure process. Remote triggering of tremor and LFEs[29], reminiscent of earthquake triggering seen in hydrothermal regions, provides additional evidence for fluid enablement.

As noted by previous authors[2,27], a slow slip event on the subduction interface would increase the stress updip on the locked portion, potentially elevating the probability of a large earthquake during or soon after the slow slip. The same concept applies to other fault settings with juxtaposed locked and creeping portions, like the San Andreas[25]. If LFEs are direct signals of episodic slow slip, they would illuminate aseismic slip at depth with temporal and spatial precision well beyond that attainable with surface geodetic instruments. With proper instrumentation and analysis, we will probably detect similar events within tremor in regions beyond southwest Japan. Consequently, LFEs may provide critical information for monitoring the seismic hazard in a variety of tectonic settings.

Received 17 January; accepted 18 May 2006.

1. Obara, K. Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* **296,** 1679–1681 (2002).
2. Rogers, G. & Dragert, H. Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* **300,** 1942–1943 (2003).
3. Dragert, H., Wang, K. & James, T. S. A silent slip event on the deeper Cascadia subduction interface. *Science* **292,** 1525–1528 (2001).
4. Miller, M. M., Melbourne, T., Johnson, D. J. & Sumner, W. Q. Periodic slow earthquakes from the Cascadia subduction zone. *Science* **295,** 2423 (2002).
5. Obara, K., Hirose, H., Yamamizu, F. & Kasahara, K. Episodic slow slip events accompanied by non-volcanic tremors in southwest Japan subduction zone. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020848 (2004).
6. Hirose, H. & Obara, K. Repeating short- and long-term slow slip events with deep tremor activity around the Bungo channel region, southwest Japan. *Earth Planets Space* **57,** 961–972 (2005).
7. Obara, K. & Hirose, H. Non-volcanic deep low-frequency tremors accompanying slow slips in the southwest Japan subduction zone. *Tectonophysics* **417,** 33–51 (2006).
8. Seno, T. & Yamasaki, T. Low-frequency tremors, intraslab and interplate earthquakes in Southwest Japan–from a viewpoint of slab dehydration. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL018349 (2003).
9. Katsumata, A. & Kamaya, N. Low-frequency continuous tremor around the Moho discontinuity away from volcanoes in the southwest Japan. *Geophys. Res. Lett.* **30**, doi:10.1029/2002GL015981 (2003).
10. Chouet, B. Resonance of a fluid-driven crack: Radiation properties and implications for the source of long-period events and harmonic tremor. *J. Geophys. Res.* **93,** 4375–4400 (1988).
11. Julian, B. R. Volcanic tremor: Nonlinear excitation by fluid flow. *J. Geophys. Res.* **99,** 11859–11878 (1994).
12. Kao, H. *et al.* A wide depth distribution of seismic tremors along the northern Cascadia margin. *Nature* **436,** 841–844 (2005).
13. Zhang, H. & Thurber, C. H. Double-difference tomography: The method and its application to the Hayward fault, California. *Bull. Seismol. Soc. Am.* **93,** 1875–1889 (2003).
14. Zhang, H. *et al.* High-resolution subducting-slab structure beneath northern Honshu, Japan, revealed by double-difference tomography. *Geology* **32,** 361–364 (2004).
15. Obara, K., Kasahara, K., Hori, S & Okada, Y. A densely distributed high-sensitivity seismograph network in Japan: Hi-net by National Research Institute for Earth Science and Disaster Prevention. *Rev. Sci. Instrum.* **76**, doi:10.1063/1.1854197 (2005).
16. Baba, T., Tanioka, Y., Cummins, P. & Uhira, K. The slip distribution of the 1946 Nankai earthquake estimated from tsunami inversion using a new plate model. *Phys. Earth Planet. Inter.* **132,** 59–73 (2002).
17. Shiomi, K., Sato, H., Obara, K. & Ohtake, M. Configuration of subducting Philippine Sea plate beneath southwest Japan revealed from receiver function analysis based on the multivariate autoregressive model. *J. Geophys. Res.* **109**, doi:10.1029/2003JB002774 (2004).
18. Peacock, S. M. & Wang, K. Seismic consequences of warm versus cool subduction metamorphism: Examples from southwest and northeast Japan. *Science* **286,** 937–939 (1999).
19. Hacker, B. R., Peacock, S. M., Abers, G. A. & Holloway, S. D. Subduction Factory 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *J. Geophys. Res.* **108**, doi:10.1029/2001JB001129 (2003).
20. Christensen, N. I. Pore pressure and oceanic crustal seismic structure. *Geophys J. R. Astron. Soc.* **79,** 411–423 (1984).
21. Raleigh, C. B. & Paterson, M. S. Experimental deformation of serpentinite and its tectonic implications. *J. Geophys. Res.* **70,** 3965–3985 (1965).
22. Kirby, S. H. Intraslab earthquakes and phase changes in subducting lithosphere. *Rev. Geophys.* **33,** 287–297 (1995).
23. Kodaira, S. *et al.* High pore fluid pressure may cause silent slip in the Nankai Trough. *Science* **304,** 1295–1298 (2004).
24. Ozawa, S. *et al.* Detection and monitoring of ongoing aseismic slip in the Tokai region, central Japan. *Science* **298**, 1009–1012 (2002); published online 3 October 2002 (doi:10.1126/science.1076780).
25. Nadeau, R. M. & Dolenc, D. Nonvolcanic tremors deep beneath the San Andreas fault. *Science* **307**, 389 (2005); published online 9 December 2004 (doi:10.1126/science.1107142).
26. Tanioka, Y. & Satake, K. Coseismic slip distribution of the 1946 Nankai earthquake and aseismic slips caused by the earthquake. *Earth Planets Space* **53,** 235–241 (2001).
27. Hirose, H., Hirahara, K., Kimata, F., Fujii, N. & Miyazaki, S. A slow thrust slip event following the two 1996 Hyuganada earthquakes beneath the Bungo Channel, southwest Japan. *Geophys. Res. Lett.* **26,** 3237–3240 (1999).
28. Miyazaki, S., McGuire, J. J. & Segall, P. A transient subduction zone slip episode in southwest Japan observed by the nationwide GPS array. *J. Geophys. Res.* **108**, doi:10.129/2001JB000456 (2003).
29. Miyazawa, M. & Mori, J. Detection of triggered deep low-frequency events from the 2003 Tokachi-oki earthquake. *Geophys. Res. Lett.* **32**. doi:10.1029/2005GL022539 (2005).
30. Bird, P. An updated digital model of plate boundaries. *Geochem. Geophys. Geosyst.* **4**, doi:10.1029/2001GC000252 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This material is based upon work supported by the National Science Foundation. We thank K. Obara for allowing us to use his determinations of daily tremor activity. Much of this work was conducted while D.R.S. was a participant in the Summer Institute in Japan programme, cosponsored by the Japan Society for Promotion of Science and the National Science Foundation. All data were obtained from the NIED Hi-net data server.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.R.S. (dshelly@pangea.stanford.edu).

LETTERS

Isolation of a novel acidiphilic methanogen from an acidic peat bog

Suzanna L. Bräuer[1], Hinsby Cadillo-Quiroz[1], Erika Yashiro[1]†, Joseph B. Yavitt[2] & Stephen H. Zinder[1]

Acidic peatlands are among the largest natural sources of atmospheric methane and harbour a large diversity of methanogenic Archaea[1]. Despite the ubiquity of methanogens in these peatlands, indigenous methanogens capable of growth at acidic pH values have resisted culture and isolation[2–4]; these recalcitrant methanogens include members of an uncultured family-level clade in the Methanomicrobiales prevalent in many acidic peat bogs in the Northern Hemisphere[1,5,6]. However, we recently succeeded in obtaining a mixed enrichment culture of a member of this clade[7]. Here we describe its isolation and initial characterization. We demonstrate that the optimum pH for methanogenesis by this organism is lower than that of any previously described methanogen.

It has been estimated that northern peatlands contain one-third of the carbon found in soils worldwide[8]. They are important sources of atmospheric methane, a greenhouse gas whose concentration has more than doubled in the past 200 years (ref. 9), reaching levels unprecedented in the past 650,000 years (ref. 10). *Sphagnum*-dominated bogs, the predominant peatlands[11], are typically acidic (pH < 5) with low concentrations of mineral nutrients. Acidiphilic aerobic methane-oxidizing organisms have been isolated[12], whereas attempts to isolate acidiphilic methanogens from bogs were unsuccessful[2,13].

16S rRNA-based culture-independent studies indicate that the methanogens present in acidic bogs comprise several novel genera and species[1,5,6,11]. A family-level clade in the H_2/CO_2-using methanoarchaeal order Methanomicrobiales—originally called the R10 group[5], and also called the fen cluster[7,14] or E1/E2 cluster[15]—is usually abundant in 16S rRNA clone libraries from acidic peat bogs in diverse geographical locations[1,5,6,16]. For example, in McLean Bog, New York State, a small acidic (pH 4.1) bog we study, 66 out of 84 methanogen 16S rRNA gene clones belonged to this clade[1]. More recent studies using quantitative polymerase chain reaction (PCR) showed nearly 10^9 cells per gram dry weight of this group in active layers of McLean Bog peat, about half of the total Archaea[15]. Recently, sequences in the R10 group were detected as minority members of low-pH methanogenic enrichment cultures from German or Siberian bogs[3,4].

We recently demonstrated that the antibiotic rifampicin allowed enrichment of H_2/CO_2-using methanogenic Archaea in McLean Bog peat, and prevented growth of H_2/CO_2-using acetogenic Bacteria[7,17]. Using results from experiments examining the physical and chemical conditions that favoured methanogenesis[17], we designed a low-ionic-strength medium, designated PM1, that allowed repeated transfer of H_2/CO_2-using methanogens at pH 5 (ref. 7). Dilutions in liquid medium to 10^{-8}, the correct order of magnitude for a dilution to extinction, produced a culture, designated 6A8, containing two morphotypes—thin straight rod cells (0.2–0.3 μm diameter, 0.8–3.0 μm long) and irregular coccoid cells (0.3–0.8 μm in diameter) (Fig. 1a, b). The coccoid cells represented 5–8% of total cells and were present in all growth phases.

We initially assumed that the coccoid cells were heterotrophic bacterial contaminants; however, no growth or methanogenesis was detected in medium amended with various organic substrates, and PCR amplifications with the universal bacterial 16S rRNA gene primers 27f and 1492r were negative. All of 39 clones amplified with primers 1Af(-1) and 1100Ar for methanogens[1,5] were a single restriction type, and a 92-clone library generated using broad-range archaeal primers (1Af and ArchLSU47) that amplified a region encoding 16S rRNA, the 16S–23S internal transcribed spacer (ITS) region, and part of the 23S rRNA (1,730 base pairs in total), yielded sequences with ≥99.7% identity, within the error rate expected for PCR amplification[18,19]. The 6A8 16S rRNA gene sequence belonged to the R10-Fen cluster-E1/E2 family, most closely related to sequences from mixed cultures (Supplementary Fig. 1).

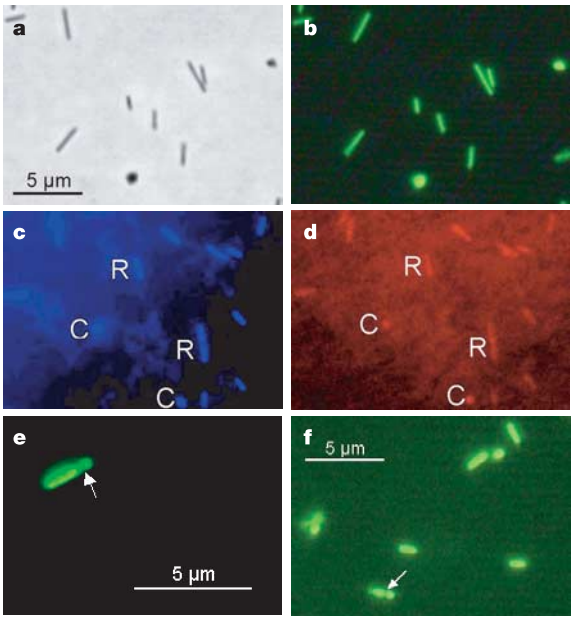


Figure 1 | Photomicrographs of culture 6A8. a, b, Phase contrast (**a**) and fluorescence (**b**) images of cells in the same field of view, stained with the DNA stain acridine orange (AO), showing thin rod and coccoid morphologies. **c, d**, Fluorescence micrographs of rod-shaped (R) and coccoid (C) cells in 6A8 stained with the DNA stain 4,6-diamidino-2-phenylindole (DAPI) (**c**) or with a Cy-3-labelled 16S rRNA probe targeting 6A8, probe 6A8 644 (**d**). **e**, AO-stained cell apparently undergoing asymmetric division to a coccoid cell (arrow). **f**, AO–stained cells passed through a 0.45-μm membrane filter and then concentrated, showing short rods and small cocci, including a cell apparently undergoing asymmetric division (arrow).

[1]Department of Microbiology and [2]Department of Natural Resources, Cornell University, Ithaca, New York 14853, USA. †Present address: Department of Plant Pathology, University of Wisconsin, Madison, Wisconsin 53706, USA.

© 2006 Nature Publishing Group

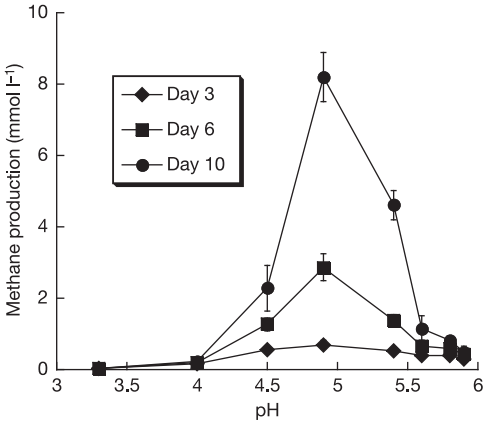


Figure 2 | Effect of pH on methanogenesis by culture 6A8 grown at 28 °C. Data points represent the averages and error bars the standard errors for triplicate samples.

Fluorescence *in situ* hybridization (FISH) staining of rRNA allows more positive identification of organisms. Initial attempts to stain these cells failed, but a FISH protocol developed for SAR11 (ref. 20), also difficult to stain, allowed visualization of both morphotypes with the universal ARCH 915 probe, but not with the EUB 338 probe (data not shown). Probe MG12006A8, in which a single-base mismatch with the 'universal' Methanomicrobiales probe MG1200 (ref. 21) was corrected, hybridized with both morphotypes (Supplementary Fig. 2), as did the strain-specific probe, 6A8 644 (Fig. 1c, d), which did not stain cells from the most closely related isolate, *Methanospirillum hungatei* JF1 (Supplementary Fig. 2).

During microscopic examination of growing cultures, cell division in rod-shaped cells was observed, but not in coccoid cells. Division in the rod-shaped cells was sometimes asymmetrical[22], giving rise to spherical cells (Fig. 1e). Filtration of cultures through a 0.45-μm filter allowed passage of short rods and small coccoid cells (Fig. 1f). In cultures grown out from this filtrate, long rods and large coccoid cells returned, and indeed Fig. 1a, b shows such a culture. From these observations, we conclude that both morphotypes are manifestations of the same organism, and suggest that the large coccoid cells are non-viable cultural artefacts.

6A8 used H_2/CO_2 as a methanogenic substrate with a doubling time near two days at pH 5.1 and 28 °C (standard growth conditions), and did not use formate, ethanol, methanol or acetate. Besides mineral nutrients, it required vitamins, yeast extract, coenzyme M and low (<200 μM) concentrations of acetate in the medium. The optimum pH for methanogenesis was near 5 (Fig. 2), with little methanogenesis in cultures incubated at pH 4.0 or 5.8. The temperature optimum was near 37 °C (Supplementary Fig. 3), with slow methanogenesis at 10 °C but not at 4 °C.

The most acidiphilic hydrogenotrophic methanogen described to date is *Methanobacterium espanolae*[23], with a pH optimum between 5.5 and 6.0 and a minimum near 4.7. Some strains of *Methanosarcina* spp. can grow near pH 4.5, although their optimum is near neutrality[24]. Microbial populations converting H_2/CO_2 to methane in McLean Bog showed an optimum pH near 5.0 (ref. 17), close to that of 6A8, but produced considerably more methane relative to the optimum at pH values near 4.3 and 5.8 than did culture 6A8, indicating that this mixed microbial population had a broader pH response. Under *in situ* conditions in McLean Bog peat (pH < 4.5, 4–20 °C), rates of methanogenesis by this organism would be considerably limited.

In summary, we have isolated a novel acidiphilic methanogen that is part of an indigenous population adapted to conditions in a methanogenic peat bog, a habitat that is not only low in pH, but also low in concentrations of mineral ions such as sodium[17]. 6A8 is clearly a member of a novel genus in the order Methanomicrobiales, and based on its preliminary characterization, we propose the name '*Candidatus* Methanoregula boonei', describing the morphology of the thin rod (from the Latin *regula*, meaning slat) and in honour of the late D. Boone, whose pioneering research on anaerobes included studies on extremophilic methanogens growing at low temperatures[25] or pH[24].

METHODS

Culture growth. Culture 6A8 was grown in anaerobic PM1 medium[7] buffered to ~pH 5.1 by the organic buffer Homopipes ($pK_a = 4.7$ at 28 °C). Methane production by cultures was quantified by gas chromatography[17].

Molecular and phylogenetic analyses. 16S rRNA gene clone libraries were generated from 6A8 DNA by PCR amplification using primers specific to methanoarchaea (1Af(-1) and 1100Ar) or Bacteria (27f and 1492r), or universal archaeal primers (1Af and ArchLSU47) spanning the 16S rRNA gene, the internal transcribed spacer (ITS) region, and part of the 23S rRNA gene. Cloned PCR amplicons were screened by restriction endonuclease digestion, and representative clones were sequenced at Cornell's Bioresource Center and were analysed using ARB and PHYLIP phylogenetic packages.

Fluorescence microscopy. Fluorescence *in situ* hybridization (FISH) was performed using fixed cells on black polycarbonate membranes[20]. The species-specific Cy3-labelled oligonucleotide probe 6A8 644 was designed using the ARB software package. Probe specificity was verified by failure to hybridize with *Escherichia coli* or *M. hungatei* JF1 cells. Stained cells were viewed using a Nikon Eclipse E600 microscope equipped with a Hamamatsu CCD digital camera, or an Olympus BX61 microscope equipped with a Cooke SensiCam camera and SlideBook software.

A more detailed description of methods is provided in Supplementary Information.

Received 19 July 2005; accepted 18 April 2006.
Published online 14 May 2006.

1. Basiliko, N., Yavitt, J. B., Dees, P. M. & Merkel, S. M. Methane biogeochemistry and methanogen communities in two northern peatland ecosystems, New York State. *Geomicrobiol. J.* **20**, 563–577 (2003).
2. Williams, R. T. & Crawford, R. L. Methanogenic bacteria including an acid-tolerant strain from peatlands. *Appl. Environ. Microbiol.* **50**, 1542–1544 (1985).
3. Sizova, M. V., Panikov, N. S., Tourova, T. P. & Flanagan, P. W. Isolation and characterization of oligotrophic acido-tolerant methanogenic consortia from a *Sphagnum* peat bog. *FEMS Microbiol. Ecol.* **45**, 301–315 (2003).
4. Horn, M. A., Matthies, C., Küsel, K., Schramm, A. & Drake, H. L. Hydrogenotrophic methanogenesis by moderately acid-tolerant methanogens of a methane-emitting acidic peat. *Appl. Environ. Microbiol.* **69**, 74–83 (2003).
5. Hales, B. A. *et al.* Isolation and identification of methanogen-specific DNA from blanket bog peat by PCR amplification and sequence analysis. *Appl. Environ. Microbiol.* **62**, 668–675 (1996).
6. Galand, P. E., Fritze, H. & Yrjala, K. Microsite-dependent changes in methanogenic populations in a boreal oligotrophic fen. *Environ. Microbiol.* **5**, 1133–1143 (2003).
7. Bräuer, S. L., Yashiro, E., Ueno, N. G., Yavitt, J. B. & Zinder, S. H. Characterization of acid-tolerant H_2/CO_2-utilizing methanogenic enrichment cultures from an acidic peat bog in New York State. *FEMS Microbiol. Ecol.* (in the press).
8. Gorham, E. Northern peatlands role in the carbon cycle and probable responses to climatic warming. *Ecol. Appl.* **1**, 182–195 (1991).
9. Cicerone, R. J. & Oremland, R. S. Biogeochemical aspects of atmospheric methane. *Glob. Biogeochem. Cycles* **2**, 299–327 (1988).
10. Brook, E. J. Atmospheric science. Tiny bubbles tell all. *Science* **310**, 1285–1287 (2005).
11. Kotsyurbenko, O. R. *et al.* Acetoclastic and hydrogenotrophic methane production and methanogenic populations in an acidic West-Siberian peat bog. *Environ. Microbiol.* **6**, 1159–1173 (2004).
12. Dedysh, S. N. *et al.* Isolation of acidophilic methane-oxidizing bacteria from northern peat wetlands. *Science* **282**, 281–283 (1998).
13. Goodwin, S. & Zeikus, J. G. Ecophysiological adaptations of anaerobic bacteria to low pH: analysis of anaerobic digestion in acidic bog sediments. *Appl. Environ. Microbiol.* **53**, 57–64 (1987).
14. Galand, P. E., Saarnio, S., Fritze, H. & Yrjala, K. Depth related diversity of methanogen Archaea in Finnish oligotrophic fen. *FEMS Microbiol. Ecol.* **42**, 441–449 (2002).
15. Cadillo-Quiroz, H. *et al.* Vertical profiles of methanogenesis and methanogens in two contrasting acidic peatlands in central New York State, USA. *Environ. Microbiol.* (in the press).
16. Juottonen, H. *et al.* Methanogen communities and Bacteria along an ecohydrological gradient in a northern raised bog complex. *Environ. Microbiol.* **7**, 1547–1557 (2005).

© 2006 Nature Publishing Group

17. Bräuer, S. L., Yavitt, J. B. & Zinder, S. H. Methanogenesis in McLean Bog, an acidic peat bog in upstate New York: stimulation by H_2/CO_2 in the presence of rifampicin, or by low concentrations of acetate. *Geomicrobiol. J.* **21,** 433–443 (2004).
18. Könneke, M. *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437,** 543–546 (2005).
19. von Wintzingerode, F., Göbel, U. B. & Stackebrandt, E. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* **21,** 213–229 (1997).
20. Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420,** 806–810 (2002).
21. Raskin, L., Stromley, J. M., Rittmann, B. E. & Stahl, D. A. Group specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl. Environ. Microbiol.* **60,** 1232–1240 (1994).
22. Angert, E. R. Alternatives to binary fission in bacteria. *Nature Rev. Microbiol.* **3,** 214–224 (2005).
23. Patel, G. B., Sprott, G. D. & Fein, J. E. Isolation and characterization of *Methanobacterium espanolae* sp. nov., a mesophilic moderately acidophilic methanogen. *Int. J. Syst. Bacteriol.* **40,** 12–18 (1990).
24. Maestrojuan, G. M. & Boone, D. R. Characterization of *Methanosarcina barkeri* MS^T and 227 *Methanosarcina mazei* S-6^T and *Methanosarcina vacuolata* Z-761^T. *Int. J. Syst. Bacteriol.* **41,** 267–274 (1991).
25. Franzmann, P. D. *et al. Methanogenium frigidum* sp. nov., a psychrophilic, H_2-using methanogen from Ace Lake, Antarctica. *Int. J. Syst. Bacteriol.* **47,** 1068–1072 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Angert and W. Ghiorse for microscopy advice and equipment, R. Morris and S. Dedysh for advice on FISH, M. McInerney for providing us with cultures of *M. hungatei* JF1, and H. Trüper for advice on naming the organism. This work was supported by the NSF Microbial Observatories Program.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.H.Z. (shz1@cornell.edu).

© 2006 Nature Publishing Group

LETTERS

A high-performance brain–computer interface

Gopal Santhanam[1]*, Stephen I. Ryu[1,2]*, Byron M. Yu[1], Afsheen Afshar[1,3] & Krishna V. Shenoy[1,4]

Recent studies have demonstrated that monkeys[1–4] and humans[5–9] can use signals from the brain to guide computer cursors. Brain–computer interfaces (BCIs) may one day assist patients suffering from neurological injury or disease, but relatively low system performance remains a major obstacle. In fact, the speed and accuracy with which keys can be selected using BCIs is still far lower than for systems relying on eye movements. This is true whether BCIs use recordings from populations of individual neurons using invasive electrode techniques[1–5,7,8] or electroencephalogram recordings using less-[6] or non-invasive[9] techniques. Here we present the design and demonstration, using electrode arrays implanted in monkey dorsal premotor cortex, of a manyfold higher performance BCI than previously reported[9,10]. These results indicate that a fast and accurate key selection system, capable of operating with a range of keyboard sizes, is possible (up to 6.5 bits per second, or ~15 words per minute, with 96 electrodes). The highest information throughput is achieved with unprecedentedly brief neural recordings, even as recording quality degrades over time. These performance results and their implications for system design should substantially increase the clinical viability of BCIs in humans.

Most BCIs translate neural activity into a continuous movement command, which guides a computer cursor to a desired visual target[1–3,5–9]. If the cursor is used to select targets representing discrete actions, the BCI serves as a communication prosthesis. Examples include typing keys on a keyboard, turning on room lights, and moving a wheelchair in specific directions. Human-operated BCIs are currently capable of communicating only a few letters per minute (~1 bits per second (bps) sustained rate[9]) and monkey-operated systems can only accurately select one target every 1–3 s (~1.6 bps sustained rate[10]), despite using invasive electrodes.

An alternative, potentially higher-performance approach is to translate neural activity into a prediction of the intended target and immediately place the cursor directly on that location. This type of control is appropriate for communication prostheses and benefits from not having to estimate unnecessary parameters such as continuous trajectory[4,11]. We conducted a series of experiments to investigate how quickly and accurately a BCI could operate under direct end-point control.

We used a standard instructed-delay behavioural task[12] to assess neural activity in the arm representation of monkey premotor cortex (PMd), as shown in Fig. 1a and described in Methods. As previously

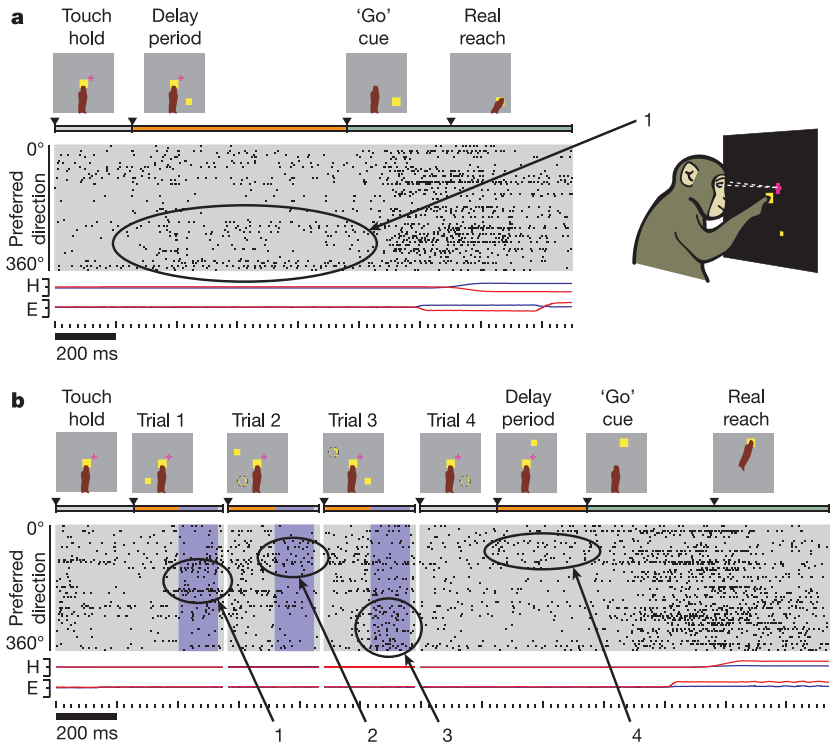


Figure 1 | Instructed-delay (real reach) and BCI (prosthetic cursor) tasks, with accompanying neural data. Large numbered ellipses draw attention to the increase in neural activity related to the peripheral reach target. **a**, Standard instructed-delay reach trial. Data from selected neural units are shown (grey shaded region); each row corresponds to one unit and black tick marks indicate spike times. Units are ordered by angular tuning direction (preferred direction) during the delay period. For hand (H) and eye (E) traces, blue and red lines show the horizontal and vertical coordinates, respectively. The full range of scale for these data is ±15 cm from the centre touch cue. **b**, Chain of three prosthetic cursor trials followed by a standard instructed-delay reach trial. T_{skip} is denoted by the orange parts of the time line. Neural activity was integrated (T_{int}) during the purple shaded interval and used to predict the reach target location. After a short processing time ($T_{dec+rend} \approx 40$ ms), a prosthetic cursor was briefly rendered and a new target was displayed. The dotted circles represent the reach target and prosthetic cursor from the previous trial, both of which were rapidly extinguished before the start of the trial indicated. Trials shown here are from experiment H20041106.1 with monkey H.

[1]Department of Electrical Engineering, Stanford University, 330 Serra Mall, 319 Paul G. Allen Center for Integrated Systems Annex, Stanford, California 94305-4075, USA. [2]Department of Neurosurgery, Stanford University School of Medicine, 300 Pasteur Drive, Edwards Building, R-297, Stanford, California 94305-5327, USA. [3]Medical Scientist Training Program, Stanford University School of Medicine, 251 Campus Drive, MSOB 309, Stanford, California 94305-5404, USA. [4]Neurosciences Program, Stanford University School of Medicine, Stanford, California 94305, USA.
*These authors contributed equally to this work.

© 2006 Nature Publishing Group

reported, neural activity during the delay period (time from target appearance until 'go' cue) reflects the end-point of the upcoming reach[13]. This was also the case if the visual target was presented only briefly, thereby confirming that sustained neural activity is motor-related rather than purely sensory (see Supplementary Information). The reach end-point can be decoded from delay-period activity using maximum-likelihood techniques[14].

Before conducting BCI experiments, we analysed the neural activity from instructed-delay control experiments to set parameters essential for high-performance BCI operation. We subdivided the delay period into two epochs: a time to skip (disregard) after target onset while waiting for reach end-point information to become reliable and therefore readily decodable (T_{skip}), and a time to integrate the neural data that will be used to predict the desired target selection (T_{int}).

The first epoch, T_{skip}, includes the time for visual information about the target to arrive in PMd (50–70 ms), the time for the subject to select among targets if more than one are present, and the time for neural activity reflecting the desired target to be generated. Despite being of considerable scientific interest[15,16], neural activity during these early periods is discarded in the present BCI design. Some activity during this period may already be predictive of the desired target, but it is not yet clear how best to decode this information. T_{skip} was chosen to be 150 ms based on control experiments including a multi-target task where the monkey was trained to reach for one of many simultaneously presented targets (see Supplementary Information).

The second epoch, T_{int}, directly follows the first and provides the neural data used to predict the desired BCI cursor position. Given the Poisson-like noise in the spike timing of cortical neurons[17] (confirmed with our data), a longer T_{int} will average away more noise and result in more accurate predictions of reach end-point. However, a longer T_{int} will also reduce the total number of cursor positionings that can be made per second. Herein lies the fundamental speed–accuracy trade-off that we must optimize in order to increase BCI performance.

To determine the best T_{int} to be used in BCI experiments, we analysed the effect of this parameter on two performance metrics. The first is single-trial accuracy, which is the percentage of targets correctly predicted. We found that accuracy rises and largely saturates around 85–90% as T_{int} increases to 200–250 ms. Figure 2a illustrates this effect as a function of total trial length, which is defined to be the sum of T_{skip} (150 ms), T_{int} (variable) and a small system overhead time associated with decoding and rendering the prosthetic cursor on the screen ($T_{dec+rend} \approx 40$ ms). Should a minimum level of single-trial accuracy be required for a particular application, a corresponding minimum T_{int} can be chosen.

The second performance metric is information transfer rate capacity (ITRC, in bps). This quantity measures the rate at which information is conveyed from the subject, through the BCI, to the environment[10,18]. It is the information per trial, which is closely related to single-trial accuracy, divided by the total trial length. As shown in Fig. 2a, the optimal ITRC occurs at short trial lengths, despite relatively low single-trial accuracy at these trial lengths. The highest ITRC is 7.7 bps at a total trial time of 260 ms, which corresponds to a T_{int} of 70 ms ($T_{skip} = 150$ ms, $T_{dec+rend} = 40$ ms). As further confirmation that neural responses are reflecting motor intention even at a rapid pace, we repeated the above analysis with the multi-target task. Despite this more-challenging task, requiring selection among many visual targets, the ITRC was diminished by a modest 30% (see Supplementary Fig. S4).

The performance curves in Fig. 2a are extrapolations using experimental data from individual trials that had long delay periods and long times between trials (Fig. 1a). To measure directly the ITRC performance when actually presenting trials at high speeds, we conducted a series of BCI experiments using a real-time system capable of rapidly decoding neural information. BCI experiments began with the collection of delay-period activity preceding reaches to different target locations (Fig. 1a) and fitting statistical models to the activity (model training). Then, during BCI prosthetic cursor trials (Fig. 1b), the intended target was decoded and a circular cursor was rendered on the screen at the predicted location. If the prediction was correct, the next target was displayed immediately. In this manner, a sequence of high-speed prosthetic cursor trials could be generated. Figure 1b illustrates three successful prosthetic cursor trials followed by a standard real reach trial.

Using this paradigm, we varied the number of locations at which a target could appear on any given trial. This allowed task difficulty to be varied, which contributes to the ITRC metric. Performance values were calculated by averaging data from several hundred trials per

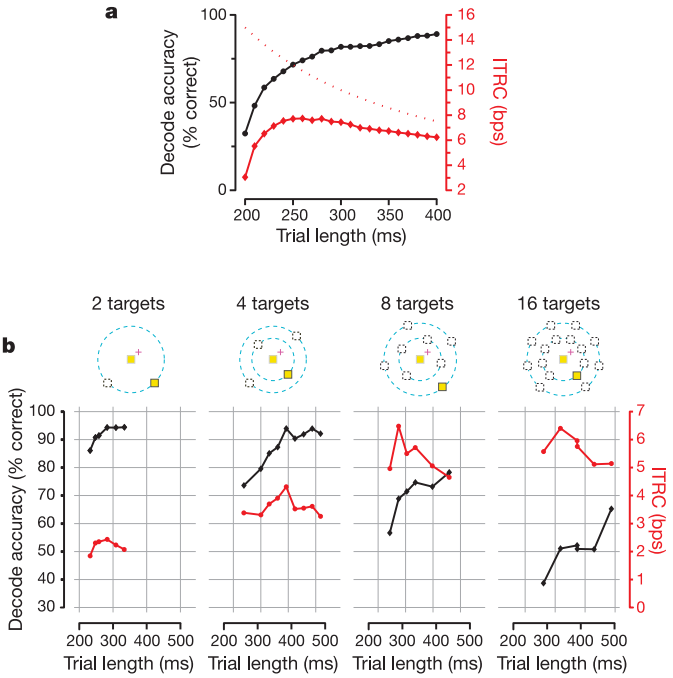


Figure 2 | Single-trial accuracy and ITRC with monkey H. a, Performance curves investigating the dependence on T_{int} were calculated from control experiment H20041118 (8-target configuration). The trial length was $T_{skip} + T_{int} + T_{dec+rend}$ with $T_{skip} = 150$ ms and $T_{dec+rend} \approx 40$ ms. T_{int} was varied and performance was computed. Performance metrics were very consistent day after day and between monkeys (data not shown). The theoretical maximum ITRC in bps, assuming 100% accuracy regardless of T_{int}, is plotted as the dotted red curve. **b**, Performance measured during BCI experiments. Performance is plotted for each target configuration and across varying total trial lengths. Each data symbol represents performance calculated from one experiment (many hundreds of trials). Across target configurations, single-trial accuracy decreases and ITRC increases as more targets locations are used.

Table 1 | BCI experiments with highest ITRC for monkeys H and G

Monkey	Number of targets	Accuracy (%)	Trials per second	bps
H	2	94.3	3.5	2.4
H	4	94.5	2.8	4.7
H	8	68.9	3.5	6.5
H	16	51.1	2.9	6.4
G	2	84.2	3.6	1.3
G	4	93.0	2.5	3.8
G	8	76.8	2.5	5.3
G	16	26.4	2.2	3.1

Each row lists the experiment with highest performance (ITRC) for a given target layout. Other experiments yielded higher single-trial accuracy or involved faster cursor rates, but did not achieve the highest ITRC for the corresponding target layout (not shown).

© 2006 Nature Publishing Group

condition. Table 1 lists the highest ITRC results during BCI experiments with 2, 4, 8 or 16 targets. The best overall performance was achieved with the 8-target task (6.5 and 5.3 bps, monkeys H and G). This performance corresponds to typing ~15 words per minute with a basic alphanumeric keyboard.

Although a sustained performance rate of 6.5 bps is manyfold greater than reported previously, it is lower than the extrapolated result (7.7 bps). Furthermore, the ITRC peak was expected at a total trial length of 260 ms, but our BCI experiments yielded 5 bps with this timing (monkey H). These discrepancies are due to the limitations inherent when using control experiments to extrapolate performance for speeds at which the subject must quickly recognize new targets and rapidly change neural activity (that is, change reach plans). The differences between extrapolated and directly measured performance were present despite specific model training methods that allowed for a fair comparison (see Supplementary Information).

Having confirmed that large BCI performance gains are possible with a direct end-point control strategy, we investigated two additional performance aspects. First, we varied T_{int} in BCI experiments with monkey H to verify experimentally the trends seen in Fig. 2a. Figure 2b also demonstrates an increase in single-trial accuracy with increasing trial length (black curves) as well as a peak in each ITRC curve (red curves). These results reveal how two or four target tasks restrict ITRC by virtue of the lower number of maximum bits per trial (1 and 2, respectively). Furthermore, given the numbers of neural units available in these experiments, it appears that ITRC is approaching a saturation point beyond which adding more target locations may not produce an appreciable increase in performance (doubling targets from 8 to 16 does not increase ITRC; although the latter layout requires distance tuning, which is known to be weaker than direction tuning[13]). Additional target locations should improve ITRC when more neurons are available.

Second, a common concern for BCIs such as ours is that as the electrode implant ages the number of recordable neurons declines, leading to a drop in overall performance[19]. To investigate the impact of neuronal loss, we performed analyses of single-trial accuracy and ITRC using data from control experiments. As expected, single-trial accuracy falls as neuron ensemble size decreases. However, it is possible to compensate partially for this performance loss by increasing T_{int}; BCI speed may be compromised as a result, but single-trial accuracy can be preserved (Supplementary Fig. S5). Figure 3 plots ITRC as a function of the number of neural units and T_{int}. For small ensembles (for example, 20 neurons), the ITRC peaks at $T_{int} \approx 120$ ms but does not decline sharply as T_{int} is further increased; accuracy (and bits per trial) is increasing so as to offset the longer trial times. For larger ensembles, the information content at small T_{int} is relatively high such that further lengthening T_{int} has a dramatic effect on ITRC. Finally, plotting the T_{int} value that maximizes ITRC (Fig. 3 inset) for each ensemble size illustrates that the maximum ITRC is achieved with small T_{int} (60–130 ms), over a broad range of ensemble sizes. Thus, high-performance BCIs may require far shorter trials than previously explored.

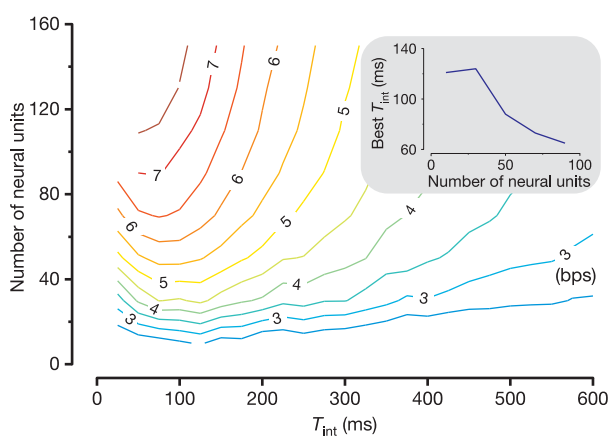


Figure 3 | ITRC as a function of number of neural units and T_{int}. All data are from experiment H20041118, which used an 8-target configuration and contained over 1,300 trials. T_{skip} was fixed at 150 ms. The main panel shows contours of ITRC (bps) as a function of the number of neural units available and T_{int}. The inset shows the value of T_{int} that achieves the maximum ITRC for each neural ensemble size that we tested. Similar results were obtained for data set G20040508 from monkey G.

Using a direct end-point control strategy, we report here a greater than fourfold (6.5 versus 1.6 bps) increase in BCI performance compared to recent studies. Performance is calculated in a conservative fashion, as the entire trial time ($T_{skip} + T_{int} + T_{dec+rend}$) was used; had just T_{int} been used, as is often done, the maximum ITRC would have been 28.4 bps, but this does not reflect an achievable selection rate. As described previously, our system differs from continuous BCI approaches in several ways, which might account for the performance gain. Additionally, continuous BCIs attempt to move the cursor well enough—although at the expense of speed (1–3 s per selection)—to avoid making errors for a given target selection. Conversely, the direct end-point control reported here need not correct errors within a given selection because these errors can be rectified with rapid follow-on selections.

Our performance results far exceed electroencephalogram (EEG)-based non-invasive system performance, and helps motivate the use of invasive, electrode-based systems in clinical BCIs. Although at its fastest, this direct end-point control BCI demonstrates selection speeds (~3.5 trials per second) on a par with saccadic eye movements, the ITRC for saccades is much higher due to their exceptional precision and accuracy. Whereas eye or even speech control may be effective in specific settings, BCIs attempting to restore lost motor function must rely on the natural neural signals if they are to avoid commandeering and interfering with another motor modality.

METHODS

See Supplementary Information for additional details on the methods of this study.

We trained two rhesus monkeys (*Macaca mulatta*) (G and H) to perform instructed-delay centre-out reaches. Animal protocols were approved by the Stanford University Institutional Animal Care and Use Committee. Hand and eye position were tracked optically (Polaris, Northern Digital; Iscan). Stimuli were back-projected onto a frontoparallel screen 30 cm from the monkey. Real reach trials (Fig. 1a) began when the monkey touched a central yellow square and fixated on a magenta cross. After a touch hold time (200–400 ms), a visual reach target appeared on the screen. After a randomized (200–1,000 ms) delay period, a 'go' cue (fixation and central touch cues were extinguished and reach target was slightly enlarged) indicated that a reach should be made to the target. Fixation was enforced throughout the delay period to control for eye-position-modulated activity in PMd[20,21]. This fixation requirement is applicable in a clinical setting if targets are near-foveal, or imagined as in a virtual keyboard set-up. The hand was also not allowed to move until the go cue was presented, providing a proxy for the cortical function of a paralysed subject[1–3]. Subsequent to a brief reaction time, the reach was executed, the target was held (~200 ms), and a juice reward was delivered along with an auditory tone. An inter-trial interval (~250 ms) was inserted before starting the next trial. We presented various target configurations (2, 4, 8 or 16 targets) on the screen, including layouts with 2, 4, or 8 directions, and 1 or 2 distances (6–12 cm radially outward).

Neural recordings. Neural activity was simultaneously recorded from a 96-channel electrode array (Cyberkinetics Neurotechnology Systems) implanted in arm representation of PMd, contralateral to the reaching arm (left, monkey G; right, monkey H). We previously reported our set-up for automatic spike sorting[22] for monkey G. A similar, but more sophisticated, system was used for monkey H. The use of an automatic spike sorting system ensured a very fast and repeatable method for classifying neural units each day. We recorded 20–30 single neurons and 60–100 multi-neuron units in a typical session.

Models and decoding. For the instructed-delay and BCI tasks, we first collected data across several hundred trials and fit (trained) models based on neural activity collected starting T_{skip} after the target presentation time and extending for a duration T_{int}. For each target location, the distribution of spike counts

© 2006 Nature Publishing Group

across trials was modelled as either a Poisson or multivariate Gaussian. We then used these models to predict the reach target on separate trials, given only the neural data from each trial.

For the instructed-delay control experiments (Figs 2a and 3; see also Supplementary Figs S2–S5), we either used two separate blocks of trials, one for training and one for prediction, or used leave-one-out cross-validation with all the trials in a data set. For BCI experiments, training trials were initially collected to fit the models and during subsequent trials, the target was predicted with the model (Gaussian model for monkey G and Poisson model for monkey H). This prediction was rendered on the screen as a circle around the predicted target location and was visible to the monkey. If the prediction was correct, an auditory tone was delivered and a subsequent target was presented with very little delay. If the prediction was incorrect, the trial was either considered a failure and aborted, or the monkey was allowed to make a real reach to the target. Real reach trials were interspersed to ensure that the monkey remained engaged in the task.

Experiments with monkey G showed consistently lower BCI performance than those with monkey H. This can be attributed to several factors, including statistical models (Gaussian versus Poisson), model training methods, and neural signal quality. When these parameters were equalized in control experiments, performance was comparable for both monkeys.

ITRC calculation. The ITRC of the system was computed using the Blahut–Arimoto algorithm[23]. This iterative algorithm produces the theoretical upper-bound on information transmission per use of a communication channel, given only the noise characteristics of the channel. In our experiments, the channel is the entire prosthetic system, with the presented target corresponding to the input and the decoded target corresponding to the output. We divide the information per trial by the average time per trial to obtain the ITRC.

Neuron dropping analysis. For a single day's experiment, we selected all neural units from the array that were responsive to target location within a desired T_{int} using an analysis of variance (ANOVA; $P < 0.05$). For each neural ensemble size of interest, our total set of neural units was subdivided by drawing 100 randomized subsets (without replacement). Performance was computed for each subset and these data were averaged within a given ensemble size. This provided a single ITRC for each T_{int} and ensemble size. We generated the contour plot (Fig. 3) by using linear interpolation across this two-dimensional surface. Furthermore, for each ensemble size tested, a cubic spline interpolation was used to estimate the particular T_{int} that maximized the ITRC (Fig. 3 inset).

Received 1 March; accepted 16 May 2006.

1. Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R. & Donoghue, J. Instant neural control of a movement signal. *Nature* **416,** 141–142 (2002).
2. Taylor, D. M., Tillery, S. I. H. & Schwartz, A. B. Direct cortical control of 3D neuroprosthetic devices. *Science* **296,** 1829–1832 (2002).
3. Carmena, J. M. *et al.* Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* **1,** E42 (2003).
4. Musallam, S., Corneil, B. D., Greger, B., Scherberger, H. & Andersen, R. A. Cognitive control signals for neural prosthetics. *Science* **305,** 258–262 (2004).
5. Kennedy, P. R., Bakay, R. A. E., Moore, M. M., Adams, K. & Goldwaithe, J. Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.* **8,** 198–202 (2000).
6. Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G. & Moran, D. W. A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.* **1,** 63–71 (2004).
7. Hochberg, L. R., Mukand, J., Polykoff, G., Friehs, G. & Donoghue, J. *Braingate Neuromotor Prosthesis: Nature and Use of Neural Control Signals. Program No. 520.17, 2005 Abstract Viewer/Itinerary Planner* (Society for Neuroscience, Washington DC, 2005).
8. Patil, P. G., Carmena, J. M., Nicolelis, M. A. L. & Turner, D. A. Ensemble recordings of human subcortical neurons as a source of motor control signals for a brain-machine interface. *Neurosurgery* **55,** 27–38 (2004).
9. Wolpaw, J. & McFarland, D. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl Acad. Sci. USA* **101,** 17849–17854 (2004).
10. Taylor, D. M., Tillery, S. I. H. & Schwartz, A. B. Information conveyed through brain-control: cursor vs. robot. *IEEE Trans. Neural Syst. Rehabil. Eng.* **11,** 195–199 (2003).
11. Shenoy, K. V. *et al.* Neural prosthetic control signals from plan activity. *Neuroreport* **14,** 591–596 (2003).
12. Cisek, P. & Kalaska, J. F. Neural correlates of mental rehearsal in dorsal premotor cortex. *Nature* **431,** 993–996 (2004).
13. Messier, J. & Kalaska, J. F. Covariation of primate dorsal premotor cell activity with direction and amplitude during a memorized-delay reaching task. *J. Neurophysiol.* **84,** 152–165 (2000).
14. Yu, B. M., Ryu, S. I., Santhanam, G., Churchland, M. M. & Shenoy, K. V. *Improving Neural Prosthetic System Performance by Combining Plan and Peri-movement Activity* 4516–4519 (Proc. 26th Annu. Int. Conf. IEEE EMBS, San Francisco, California, 2004).
15. Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G. & Shenoy, K. V. Neural variability in premotor cortex provides a signature of motor preparation. *J. Neurosci.* **26,** 3697–3712 (2006).
16. Yu, B. M. *et al.* Extracting dynamical structure embedded in neural activity. In *Advances in Neural Information Processing Systems 18* (eds Weiss, Y., Schölkopf, B. & Platt, J.) (MIT Press, Cambridge, Massachusetts, 2006).
17. Dayan, P. & Abbott, L. F. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, Massachusetts, 2001).
18. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27,** 379–423 and 623–656 (1948).
19. Polikov, V. S., Tresco, P. A. & Reichert, W. M. Response of brain tissue to chronically implanted neural electrodes. *J. Neurosci. Methods* **148,** 1–18 (2005).
20. Cisek, P. & Kalaska, J. F. Modest gaze-related discharge modulation in monkey dorsal premotor cortex during a reaching task performed with free fixation. *J. Neurophysiol.* **88,** 1064–1072 (2002).
21. Batista, A. *et al. Heterogeneous Coordinate Frames for Reaching in Macaque PMd. Program No. 363.12, 2005 Abstract Viewer/Itinerary Planner* (Society for Neuroscience, Washington DC, 2005).
22. Santhanam, G., Sahani, M., Ryu, S. I. & Shenoy, K. V. *An Extensible Infrastructure for Fully Automated Spike Sorting During Online Experiments* 4380–4384 (Proc. 26th Annu. Int. Conf. IEEE EMBS, San Francisco, California, 2004).
23. Cover, T. & Thomas, J. *Elements of Information Theory* (John Wiley, New York, 1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Eisensee for administrative assistance, M. Howard for surgical assistance and veterinary care, and N. Hatsopoulos for surgical assistance (monkey G implant). We also thank M. Churchland and M. Sahani for scientific discussions; E. Knudsen and T. Moore for comments on the manuscript; and A. Batista for collecting the blink-memory data. This study was supported by NDSEG Fellowships (G.S. and B.M.Y.), NSF Graduate Research Fellowships (G.S. and B.M.Y.), the Christopher Reeve Foundation (S.I.R. and K.V.S.), Bio-X Fellowship (A.A.), the NIH Medical Scientist Training Program (A.A.), and the following awards to K.V.S.: a Burroughs Wellcome Fund Career Award in the Biomedical Sciences, the Stanford Center for Integrated Systems, the NSF Center for Neuromorphic Systems Engineering at Caltech, ONR Adaptive Neural Systems, the Sloan Foundation, and the Whitaker Foundation.

Author Contributions G.S. and S.I.R. contributed equally to this work. S.I.R. was responsible for the initial experimental concept and surgical implantation; G.S. and S.I.R. were responsible for experimental design, animal training, data collection and preliminary analysis; G.S. was responsible for infrastructure development, in-depth analysis and writing of the paper. B.M.Y. and A.A. participated in animal training, data collection and analysis. K.V.S. was involved in all aspects of the study.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.V.S. (shenoy@stanford.edu).

© 2006 Nature Publishing Group

LETTERS

A germline-specific class of small RNAs binds mammalian Piwi proteins

Angélique Girard[1,2], Ravi Sachidanandam[1], Gregory J. Hannon[1] & Michelle A. Carmell[1]

Small RNAs associate with Argonaute proteins and serve as sequence-specific guides to regulate messenger RNA stability, protein synthesis, chromatin organization and genome structure[1–3]. In animals, Argonaute proteins segregate into two subfamilies[4]. The Argonaute subfamily acts in RNA interference and in microRNA-mediated gene regulation using 21–22-nucleotide RNAs as guides. The Piwi subfamily is involved in germline-specific events such as germline stem cell maintenance and meiosis. However, neither the biochemical function of Piwi proteins nor the nature of their small RNA guides is known. Here we show that MIWI, a murine Piwi protein, binds a previously uncharacterized class of ~29–30-nucleotide RNAs that are highly abundant in testes. We have therefore named these Piwi-interacting RNAs (piRNAs). piRNAs show distinctive localization patterns in the genome, being predominantly grouped into 20–90-kilobase clusters, wherein long stretches of small RNAs are derived from only one strand. Similar piRNAs are also found in human and rat, with major clusters occurring in syntenic locations. Although their function must still be resolved, the abundance of piRNAs in germline cells and the male sterility of *Miwi* mutants suggest a role in gametogenesis.

In multiple organisms, mutations in *Piwi*-family genes cause defects in germline development[5]. In flies, *piwi* mutations have both cell autonomous and non-autonomous effects on female germline stem cells[6,7]. In addition, both Piwi and Aubergine are important for repression of repetitive elements[8–15]. In mammals, expression of the three Piwi subfamily members, MIWI, MILI and MIWI2, is mainly germline-restricted. Moreover, mice bearing targeted mutations in *Miwi* (ref. 16), *Mili* (ref. 17) and *Miwi2* (M.A.C., A.G. and G.J.H., unpublished data) are male-sterile with distinct defects in spermatogenesis. Therefore, we searched for potential small RNA guides for mammalian Piwi proteins in the male germ line. As a first step, we surveyed small RNA populations in testes by pCp labelling (not shown). In addition to a 21–22-nucleotide species, a species of 30 nucleotides was strongly labelled. In parallel, we prepared total RNA from several mouse tissues and visualized small RNAs by staining with SYBR gold. Although this technique could barely detect microRNAs (miRNAs), a prominently staining 30-nucleotide RNA species was visible specifically in testis RNA (Fig. 1a and Supplementary Fig. S1). These small RNAs were cloned using established procedures[18], and three sequenced RNAs, piR-1, piR-2 and piR-3, were used in northern blot analyses to confirm tissue specificity (Fig. 1b–d).

As these RNAs were abundant in testes, they represented possible small RNA partners for Piwi proteins. We therefore examined MIWI immunoprecipitates for their presence. Northern blotting revealed that the ~30-nucleotide piR-1 RNA species specifically co-purified with MIWI but not with Ago2. Similar results have been found with other randomly selected piRNAs (not shown). Interestingly, a ~22-nucleotide version of piR-1 was observed under some conditions, but the generality of such variants is unclear (not shown). Probing the same immunoprecipitates for let-7c showed a strong and specific signal in Ago2 complexes, but minimal signal in MIWI immunoprecipitates. Although MIWI may also function together with miRNAs, our results indicate that its major binding partner, and perhaps that of other Piwi family members, is a previously undescribed class of ~30-nucleotide small RNAs, which we have termed Piwi-interacting RNAs (piRNAs).

We obtained a more detailed picture of this new small RNA class using a highly parallel pyrosequencing methodology[19]. Some 87,463 reads were obtained, 52,934 of which are classified as candidate piRNAs (see Supplementary Information). Of these, 26% were cloned only once, indicating that the piRNA population in testes is complex but that sufficient sequencing depth was achieved to obtain a representative sample. One notable feature was an overwhelming bias for uracil at the first position (Supplementary Fig. S2B). Overall, 94.2% of sequences began with U. Remaining positions did not show any dramatic biases. Although the underlying cause of this unusual pattern is unknown, a less extreme 5′ nucleotide bias has been noted in other classes of small RNAs, including repeat-associated siRNAs (rasiRNAs), miRNAs and effective short-interfering RNAs (siRNAs)[20].

piRNA distribution among chromosomes did not correlate either with gene density or repeat density. When normalized to chromosome length, some chromosomes were piRNA-rich whereas others were piRNA-poor (Supplementary Fig. S2C). Most piRNAs (83.7%)

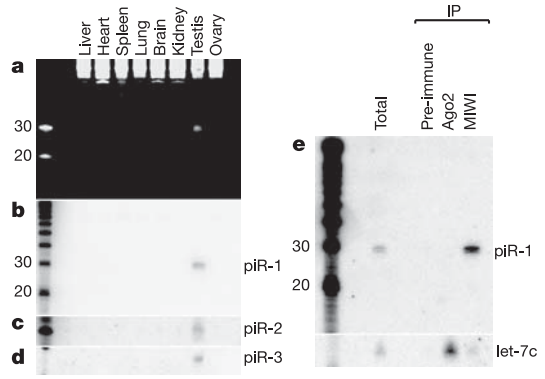


Figure 1 | Novel testis-specific small RNAs associate with MIWI. a, SYBR gold staining of 20 μg of total RNA from various mouse tissues. **b–d**, Northern blot analyses using three cloned sequences: piR-1 (**b**), piR-2 (**c**) and piR-3 (**d**). **e**, Immunoprecipitates from testis extracts with antibodies directed against Argonaute2 and MIWI were analysed by northern blotting, using piR-1 and let-7c as probes. Size (nucleotides) is shown along the left of each panel.

[1]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [2]Ecole Nationale Supérieure des Mines de Paris, 60 boulevard Saint-Michel, 75272 Paris, France.

© 2006 Nature Publishing Group

mapped uniquely in the genome (Supplementary Fig. S2D). Seventeen per cent mapped to repeats, including short interspersed elements (SINEs), long interspersed elements (LINEs), long terminal repeat (LTR) retrotransposons, and a variety of other repeat classes (Supplementary Table S1). Very few mapped to major and minor satellites that comprise the mouse centromere and none mapped to telomeric repeats. Although repeat-associated piRNAs form a significant portion of the total, this fraction is much less than would be expected based on a random distribution. Thus, mammalian piRNAs are not likely to be functional equivalents of *Drosophila* rasiRNAs, which are ~23–26 nucleotides in length and mostly derived from various classes of transposons. However, resolution of this issue will await a fuller understanding of both classes of RNAs and the identification of an Argonaute binding partner for rasiRNAs.

An examination of the 52,329 piRNAs that map to between one and five discrete loci revealed a non-uniform distribution in the genome. This analysis showed clustering of approximately 50,331 (96.2%) piRNAs into only a relatively small number of loci (Supplementary Table S2). Overall, most chromosomes contained identifiable clusters (Fig. 2a and Supplementary Tables S2 and S3). At the depth of our present study, we found that the major clusters range in size from 10 to 83 kilobases and contain between 10 and 4,500 small RNAs. They generally occur in gene- and repeat-poor regions, although rare expressed sequence tags (ESTs) from these loci were often derived from testis libraries.

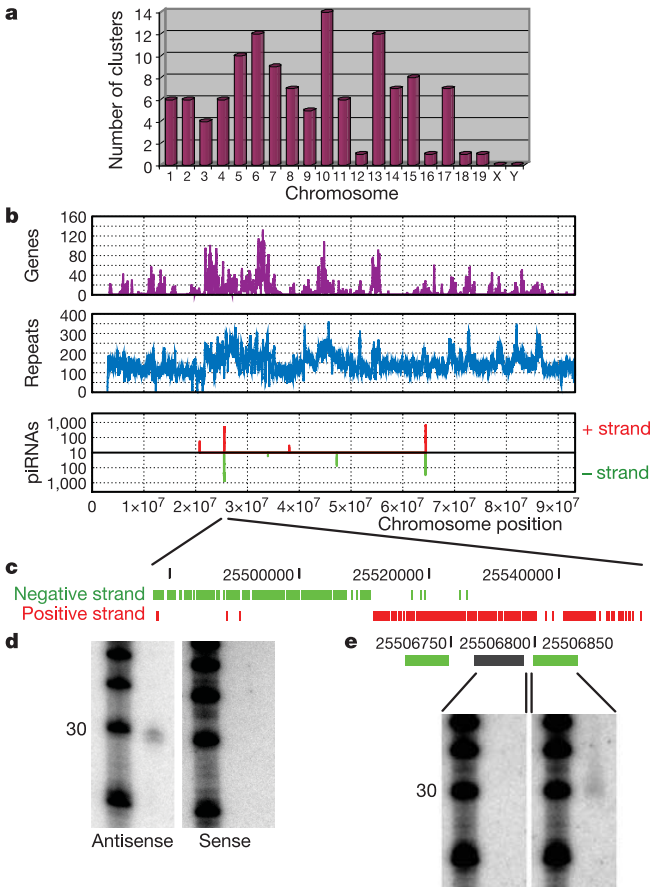


Figure 2 | Properties of the piRNA clusters. a, Absolute number of piRNA clusters per chromosome. **b**, Density analysis of exons (top), repeats (middle) and piRNAs (bottom, with positive strand in red and negative strand in green) along chromosome 17. Peaks correspond to clusters of 10 or more piRNAs. **c**, Strand bias of piRNAs. **d**, Strand specificity of piRNAs was analysed by northern blotting, using piR-1 sense and antisense probes. **e**, Non-uniform distribution of piRNAs across clusters shown by northern blotting using the frequently cloned piR-3 (green), and an uncloned, immediately adjacent region (black). Size (nucleotides) is shown along the left of panels **d** and **e**.

As one example, chromosome 17 contains 3,885 clustered piRNAs as compared to 111 piRNAs that map as non-clustered orphans (Fig. 2b). A detailed view of one large cluster revealed an asymmetric distribution of piRNAs along each genomic strand (Fig. 2c). The cluster began proximal to the centromere with a 43,400-base-pair stretch containing 1,371 piRNAs along the upper strand and then switched abruptly to a 49,300-base-pair stretch containing 1,290 piRNAs along the lower strand. Additional bi-directional clusters were found, although the two arms were not always of equal length. Other clusters were found only on one strand (unidirectional, see Supplementary Fig. S3 and Supplementary Table S2). Indeed, virtually all piRNA clusters showed a profound strand asymmetry. Density maps of piRNAs along the remaining mouse chromosomes can be found in Supplementary Figs S3–S11.

Mature piRNAs were single-stranded, as verified by northern blotting for each strand of a selected piRNA (Fig. 2d). When piRNAs were detected from protein-coding genes, they almost invariably arose only from the sense strand of the mRNA. The piRNA-generating loci themselves were not obviously enriched for secondary structure. Although some clusters did arise from a complex series of direct or inverted repeats, these were in the minority. Within any given cluster, there appeared to be gaps in piRNA production. This was confirmed by probing northern blots with immediately adjacent sequences, one of which matched a piRNA recovered in our sequence analysis (Fig. 2e). Such gaps could reflect the fact that different piRNAs show substantial variation in abundance, with low-abundance piRNAs being absent from our sequencing and not detected by northern blotting. Considered together, our observations seem to be at odds with conventional mechanisms for producing small RNAs through Dicer- and Drosha-mediated cleavage of double-stranded RNA precursors. However, there are examples, such as target-dependent accumulation of siRNAs in *Caenorhabditis elegans*, where apparently conventional small RNA generation mechanisms can give rise to a severely asymmetric population (Grishok, A. and Mello, C. C., personal communication[21]).

By SYBR gold staining, we also detected potential piRNA populations in testes of other mammals, including human, rat and bull (not shown). However, only a small number of mouse piRNAs could be mapped to the genomes of other species. We therefore sequenced candidate piRNA populations from human and rat testes, obtaining 52,099 and 47,024 sequences, respectively. The human and rat piRNAs have the same defining characteristics as mouse piRNAs, being roughly 29–30 nucleotides in length, having a strong bias for U in the first nucleotide, and clustering in their respective genomes. Despite a lack of primary sequence conservation, bi-directional clusters of piRNAs were found on human chromosome 6 and rat chromosome 20 at loci that correspond to the mouse chromosome 17 cluster shown in Fig. 2c (Fig. 3a). Indeed, the vast majority of large mouse piRNA clusters existed at syntenic loci in human and rat, whereas smaller clusters were much less likely to be found at corresponding chromosomal positions in other species (Fig. 3b and Supplementary Tables S3 and S4).

The mouse testis contains several distinct cell types including germ cells, Sertoli cells and Leydig cells. Northern blotting indicated that piRNAs are present in germ cells, as piRNAs were absent from a *c-kit* mutant (W/Wv) that lacks germ cells and from a Sertoli cell line (TM4) (Fig. 4a). These RNAs were also lacking from the epididymis, suggesting an absence from maturing and mature sperm. The first wave of spermatogenesis initiates shortly after birth and occurs fairly synchronously throughout the organ. Therefore, distinct cell types appear at defined times postnatally. Thus, the timing of the appearance of piRNAs provides clues to the stages of spermatogenesis in which they might act (Fig. 4b). piRNAs first became visible at 14 days post partum, which corresponds to the time when cells enter the pachytene stage of meiosis. Abundance increased steadily to reach

© 2006 Nature Publishing Group

adult levels at day 17–18. Overall, these data indicate that piRNAs are abundant in cells beginning at prophase of meiosis I and are lost at some point before the production of mature sperm. This corresponds to the time when all three members of the Piwi family in mouse are expressed and the time during gametogenesis when effects of mutations in these genes are apparent.

Cell-based and biochemical systems have given us an increasingly deep understanding of the mechanisms by which Argonaute family proteins use small RNAs to enforce gene silencing. However, the biochemical function of an entire clade of Argonaute proteins, the Piwi proteins, has remained a mystery, partially because their expression is largely restricted to cells undergoing gametogenesis, a process that is difficult to reproduce *ex vivo*. Here, we provide biochemical evidence for a new class of small RNAs that act as binding partners for the Piwi family. Although this represents an important step in illuminating Piwi function, many questions remain, particularly the biochemical function of the Piwi–piRNA complex. By analogy to the better understood Argonaute proteins, piRNAs are probably used to recognize substrates through base-pairing interactions. Given genetic analysis of the Piwi family and the timing of piRNA accumulation, it is tempting to imagine that these complexes perform essential functions during gametogenesis. For example, localized production of large quantities of unique RNAs could facilitate homology searching and chromosome pairing. If Piwi–piRNA complexes function in this manner, then either piRNAs must exist at some level in the ovary or there must be marked differences between male and female meioses. The piRNAs might alternatively be involved in regulating gene expression profiles or genome structure during gametogenesis. Development from primordial germ cells to mature sperm represents a period of substantial

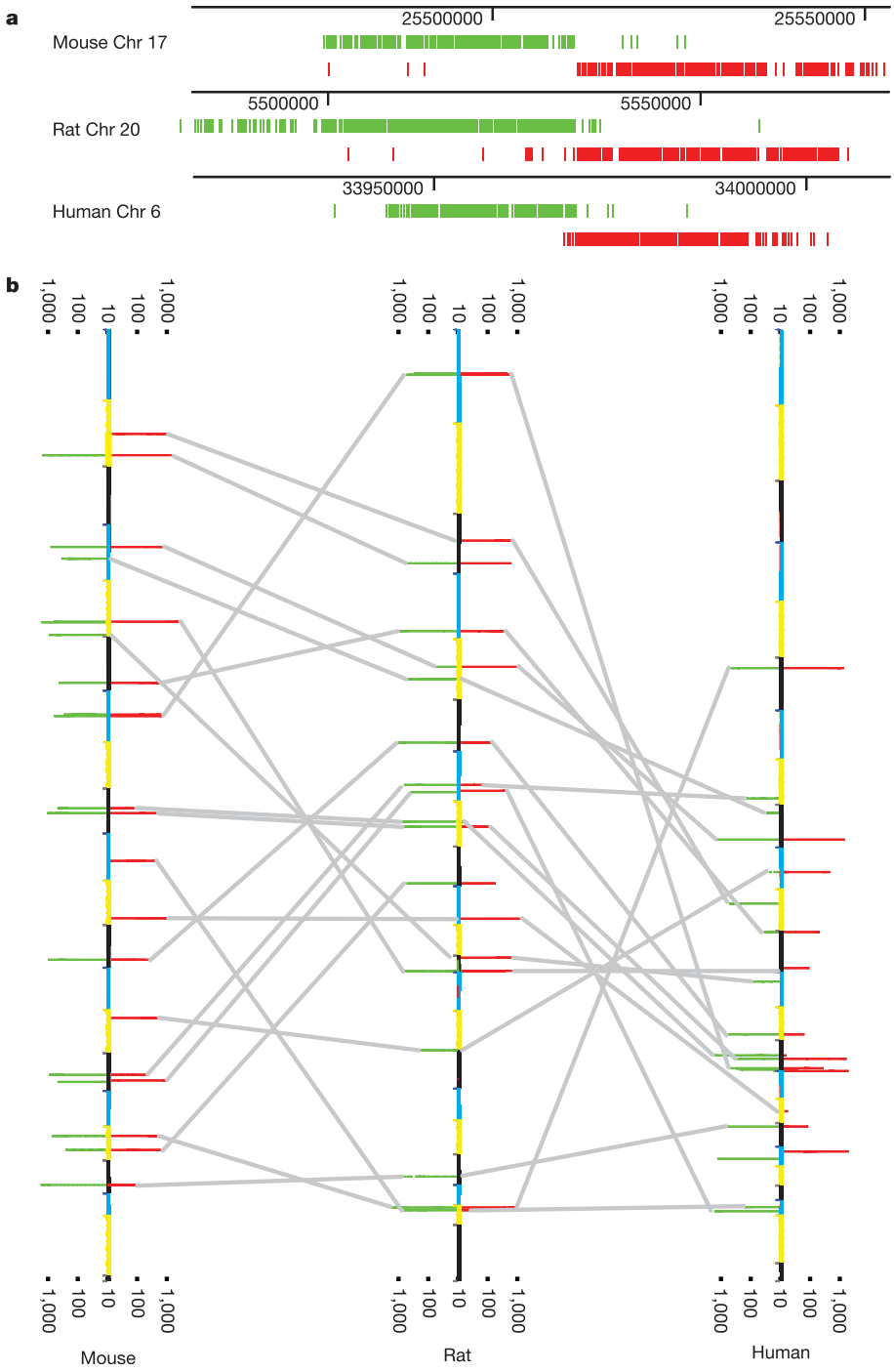


Figure 3 | Conservation of piRNA clusters among mammalian species. **a**, piRNAs cloned from a region of synteny between mouse, rat and human. Red, positive strand; green, negative strand. **b**, Genome-wide representation of the synteny of clusters. The entire genome of each species is represented vertically with chromosome 1 at the top and the sex chromosomes at the bottom. Grey lines indicate synteny. Only clusters consisting of more than 1,000 piRNAs, or showing synteny to clusters of more than 1,000 piRNAs, are depicted.

© 2006 Nature Publishing Group

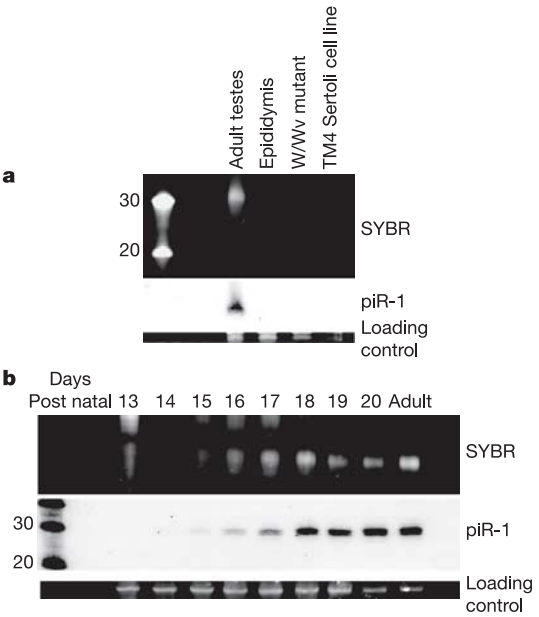


Figure 4 | piRNAs are germ cell specific and developmentally regulated. a, Germ cell specificity of the piRNAs. We examined the presence of piRNAs in somatic cells and sperm by SYBR gold staining and northern blot analysis using piR-1. **b,** Developmentally timed appearance of piRNAs. Testes were harvested during the first wave of spermatogenesis and assayed for the presence of piRNAs by SYBR gold staining and northern blotting using piR-1. Size (nucleotides) is shown along the left of each panel.

epigenetic fluidity, and mutations that disrupt the ability of cells to carry out epigenetic programming are known to have meiotic phenotypes[22–24]. Finally, spermatogenesis involves dynamic regulation of both global protein synthesis and translation of individual mRNAs, and analogies to miRNA-mediated repression suggest possible roles for piRNAs in these processes. Comparative genomics has revealed conservation of the piRNA-producing capacity of a locus, rather than conservation of the piRNAs themselves. This may indicate that the key to the function of piRNAs may not be in their precise sequence. Instead, their function may stem from the fact that they are produced in large numbers from unique loci in the genome. Genetic studies of these complex RNA populations, combined with biochemical characterization of the complexes that contain them, will ultimately be required to fit this new addition to the small RNA world into our evolving picture of the biological roles of the Argonaute gene family.

METHODS

Total RNA was prepared using Trizol (mouse, rat) or purchased from Ambion (human). Small RNAs were separated by 15% denaturing polyacrylamide gel electrophoresis and stained with SYBR gold. Immunoprecipitations were carried out using a rabbit polyclonal antibody to MIWI after lysis of testis in RIPA or Triton buffer. Small RNA northern blots were carried out essentially as described previously[25]. Probe sequences can be found in Supplementary Methods. Small RNA cloning was done as in ref. 18, and large-scale sequencing done as recommended by 454 Life Sciences. Detailed protocols and bioinformatics methods are available as Supplementary Information.

Received 18 February; accepted 18 May 2006.
Published online 4 June 2006.

1. Hannon, G. J. RNA interference. *Nature* **418,** 244–251 (2002).
2. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431,** 343–349 (2004).
3. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116,** 281–297 (2004).
4. Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16,** 2733–2742 (2002).
5. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev.* **12,** 3715–3727 (1998).
6. Lin, H. & Spradling, A. C. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* **124,** 2463–2476 (1997).
7. Cox, D. N., Chao, A. & Lin, H. *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127,** 503–514 (2000).
8. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9,** 315–327 (2002).
9. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303,** 669–672 (2004).
10. Aravin, A. A. *et al.* Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. *Mol. Cell. Biol.* **24,** 6742–6750 (2004).
11. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11,** 1017–1027 (2001).
12. Kalmykova, A. I., Klenov, M. S. & Gvozdev, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res.* **33,** 2052–2059 (2005).
13. Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A. & Gvozdev, V. Telomere elongation is under the control of the RNAi-based mechanism in the *Drosophila* germline. *Genes Dev.* **20,** 345–354 (2006).
14. Vagin, V. V. *et al.* The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*. *RNA Biol.* **1,** 51–55 (2004).
15. Sarot, E., Payen-Groschene, G., Bucheton, A. & Pelisson, A. Evidence for a *piwi*-dependent RNA silencing of the *gypsy* endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. *Genetics* **166,** 1313–1321 (2004).
16. Deng, W. & Lin, H. *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* **2,** 819–830 (2002).
17. Kuramochi-Miyagawa, S. *et al. Mili*, a mammalian member of *piwi* family gene, is essential for spermatogenesis. *Development* **131,** 839–849 (2004).
18. Ambros, V. & Lee, R. C. Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. *Methods Mol. Biol.* **265,** 131–158 (2004).
19. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
20. Aravin, A. A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5,** 337–350 (2003).
21. Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107,** 465–476 (2001).
22. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431,** 96–99 (2004).
23. Webster, K. E. *et al.* Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc. Natl Acad. Sci. USA* **102,** 4068–4073 (2005).
24. Peters, A. H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107,** 323–337 (2001).
25. Caudy, A. A., Myers, M., Hannon, G. J. & Hammond, S. M. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.* **16,** 2491–2496 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Sheth for help with the informatics analysis of piRNA populations, and W. R. McCombie for pilot sequencing. C. Cordon-Cardo and M. Golding provided tissue samples. We thank C. Mello, L. Murchison, S. Cheloufi and O. Tam for helpful discussions. This work was supported in part by grants from the NIH (G.J.H.). A.G. is a Florence Gould Fellow of the Watson School of Biological Sciences. G.J.H. is an Investigator of the Howard Hughes Medical Institute.

Author Information The GenBank accession numbers for piR-1, piR-2 and piR-3 are DQ539889, DQ539890 and DQ539891, respectively. Mouse piRNA accession numbers range from DQ539889 to DQ569912; human piRNA accession numbers range from DQ569913 to DQ601958; and rat piRNA accession numbers range from DQ601959 to DQ628526. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.edu).

© 2006 Nature Publishing Group

LETTERS

A novel class of small RNAs bind to MILI protein in mouse testes

Alexei Aravin[1*†], Dimos Gaidatzis[2*], Sébastien Pfeffer[1†], Mariana Lagos-Quintana[1], Pablo Landgraf[1], Nicola Iovino[1], Patricia Morris[3], Michael J. Brownstein[4], Satomi Kuramochi-Miyagawa[5], Toru Nakano[5], Minchen Chien[6], James J. Russo[6], Jingyue Ju[6,7], Robert Sheridan[8], Chris Sander[8], Mihaela Zavolan[2] & Thomas Tuschl[1]

Small RNAs bound to Argonaute proteins recognize partially or fully complementary nucleic acid targets in diverse gene-silencing processes[1–4]. A subgroup of the Argonaute proteins—known as the 'Piwi family'[5]—is required for germ- and stem-cell development in invertebrates[6,7], and two Piwi members—MILI and MIWI—are essential for spermatogenesis in mouse[8,9]. Here we describe a new class of small RNAs that bind to MILI in mouse male germ cells, where they accumulate at the onset of meiosis. The sequences of the over 1,000 identified unique molecules share a strong preference for a 5′ uridine, but otherwise cannot be readily classified into sequence families. Genomic mapping of these small RNAs reveals a limited number of clusters, suggesting that these RNAs are processed from long primary transcripts. The small RNAs are 26–31 nucleotides (nt) in length—clearly distinct from the 21–23 nt of microRNAs (miRNAs) or short interfering RNAs (siRNAs)—and we refer to them as 'Piwi-interacting RNAs' or piRNAs. Orthologous human chromosomal regions also give rise to small RNAs with the characteristics of piRNAs, but the cloned sequences are distinct. The identification of this new class of small RNAs provides an important starting point to determine the molecular function of Piwi proteins in mammalian spermatogenesis.

We immunoprecipitated MILI-containing ribonucleoprotein complexes from testis lysate of adult mice and purified the associated RNAs. Although 5′ ^{32}P-labelling of the isolated molecules revealed a distinct population of RNAs 26–28 nt in length, total testis RNA from fertile adults—but not 10-day-old mice—showed a prominent RNA species of approximately 30 nt in length (Fig. 1a). Labelling of the RNA remaining in the supernatant of the MILI immunoprecipitation showed that the 30-nt fraction was intact, and that the 26–28-nt MILI-interacting RNAs were unlikely to represent degradation products of the 30-nt-long RNAs. To determine the identity of the different small-RNA size populations, we cloned and sequenced the MILI-interacting small RNAs, as well as small RNAs from testis ranging in size between 18–26 nt and 24–33 nt.

The over 15,000 sequences identified in three small-RNA libraries suggested three distinct size populations of small RNAs. The 18–26-nt-fractionated library showed a peak at 21 nt and 22 nt, corresponding to miRNAs, and also revealed a 26-nt shoulder (Fig. 1b). The 24–33-nt fraction showed a bimodal distribution with a strong peak at 29–31 nt, corresponding to the small RNAs detected by 5′-labelling of total testis RNA, and a small peak at 26–28 nt. MILI-interacting small RNAs demonstrated a unimodal length-distribution, with a peak at 26–28 nt. The 27-nt shoulders in the size distribution profiles of the smaller and larger size fractions thus probably represent the MILI-interacting subpopulation. Whereas ~60% of small RNA clones from the 18–26-nt library can be annotated as miRNAs and degradation products of abundant cellular RNAs, more than 80% of the sequences in the MILI immunoprecipitate and the 24–33-nt libraries did not derive from known transcripts or genomic repeats (Supplementary Table 1).

Sequence analysis indicated that 85% of the MILI-interacting RNAs and 88% of the 24–33-nt RNAs contain a 5′ uridine residue (Supplementary Table 1, Supplementary Fig. 1). The bias for 5′ uridine is one of the characteristics of miRNAs and repeat-associated siRNAs (rasiRNAs) produced from double-stranded RNA (dsRNA) precursors by RNase III enzymes[10–12]. The majority of mouse small RNA sequences (92.9% and 97.7%, respectively) obtained from libraries of 24–33-nt testis and MILI immunoprecipitation were cloned only once. Genomic mapping of cloned sequences showed that less than 15% of clones in both libraries are derived from annotated repetitive regions (Supplementary Tables 1 and 2). Further more, less than 3% of the sequences match the genome more than ten times. The relative frequency of sequences matching different types of repeats roughly corresponds to the relative proportion of those repeat elements in the genome.

Clustering the genomic loci corresponding to the more than 1,500 MILI-interacting sequences, on the basis of a maximum distance of 15 kilobases (kb) between two consecutive loci, indicated that 81% of clones are derived from only 42 genomic regions (Fig. 2a; Supplementary Tables 3 and 4). Moreover, 19% of the sequences in the 18–26-nt library also originate in these regions, indicating that the 26-nt shoulder in the size distribution profile of this library indeed corresponds to MILI-interacting RNAs. Even more surprisingly, we found that 76% of all sequences identified in the 24–33-nt library map to the same regions, which thus seem to produce both the 26–28-nt MILI-interacting RNAs and the abundant 29–31-nt RNAs that are present in total testis RNA. On the basis of the interaction of these small RNAs with the MILI member of the Piwi protein family, we refer to them as 'Piwi-interacting RNAs' (piRNAs).

[1]Howard Hughes Medical Institute, Laboratory of RNA Molecular Biology, The Rockefeller University, 1230 York Avenue, Box 186, New York, New York 10021, USA. [2]Biozentrum, Universität Basel, Klingelbergstr 50-70, CH-4056 Basel, Switzerland. [3]Population Council, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA. [4]J. Craig Venter Institute, Functional Genomics, 9704 Medical Center Drive, Rockville, Maryland 20850, USA. [5]Department of Pathology, Medical School, Graduate School of Frontier Biosciences, Osaka University, Yamada-oka 2-2 Suita, Osaka 565-0871, Japan. [6]Columbia Genome Center, Russ Berrie Pavilion, 1150 St. Nicholas Avenue, New York, New York 10032, USA. [7]Department of Chemical Engineering, Columbia University, 500 West 120 Street, New York, New York 10027, USA. [8]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. †Present addresses: Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA (A.A.); CNRS-UPR 2357, IBMP, 12 rue du Général Zimmer, 67084 Strasbourg Cedex, France (S.P.).
*These authors contributed equally to this work.

© 2006 Nature Publishing Group

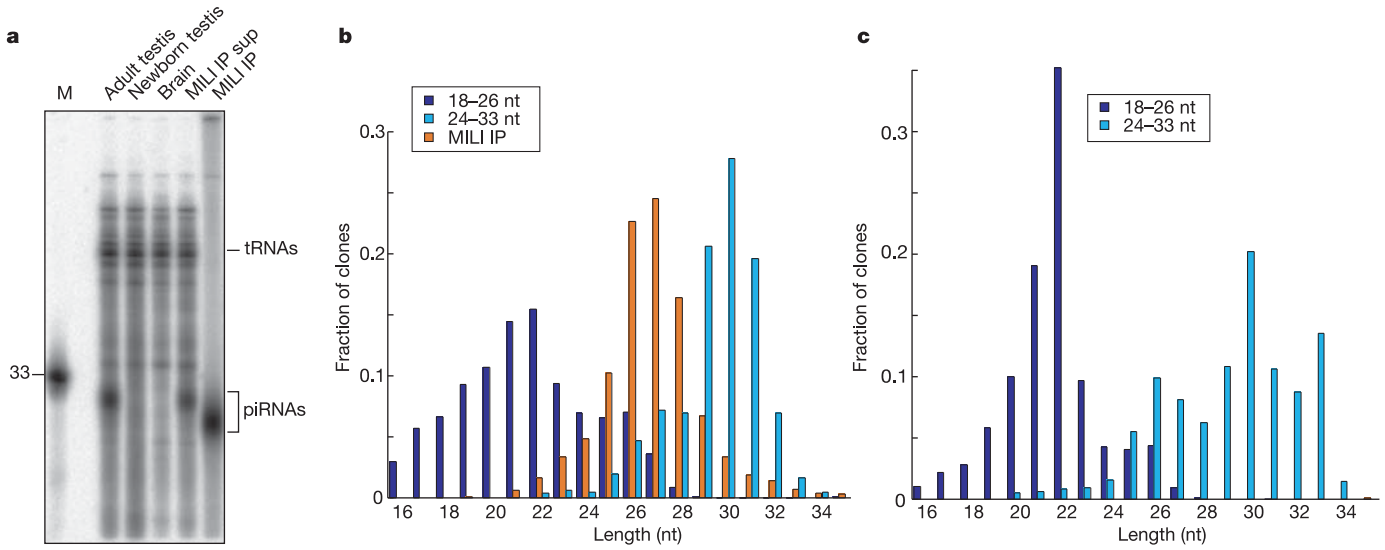


Figure 1 | MILI interacts with 26–28-nt RNAs. a, Radioactive labelling of total RNA isolated from adult mouse testis or 10-day-old newborn mouse testis, and adult mouse brain. Adult testis shows an abundant 29–31-nt small-RNA fraction. MILI-immunoprecipitated (IP) RNAs are predominantly 26–28 nt long, whereas 29–31-nt-long RNAs remain in the supernatant (sup). The size and mobility of the oligoribonucleotide marker (M) is indicated on the left; the mobility of transfer RNAs (tRNAs) and 'Piwi- interacting RNAs' (piRNAs) is indicated on the right. **b**, **c**, The size distribution (in nucleotides (nt)) of small RNAs cloned from adult mouse (**b**) and human (**c**) testis total RNA in the 18–26-nt (dark blue bars), 24–33-nt (light blue bars), and MILI-immunoprecipitated (orange bars) RNA fractions.

The pronounced clustering strongly suggests that multiple piRNAs are processed from long primary transcripts—a hypothesis supported by testis-specific expressed sequence tags (ESTs) and messenger RNAs mapping to these regions. The identified piRNA-encoding regions are distributed over most mouse chromosomes and they range in size from 0.9 to 127 kb (Fig. 2a; Supplementary Tables 3 and 4). Although piRNAs map exclusively to one chromosomal strand in many regions, some regions—including the region on mouse chromosome 17 that is one of the major sources of piRNAs—encode piRNAs in both orientations. Notably, in all cases

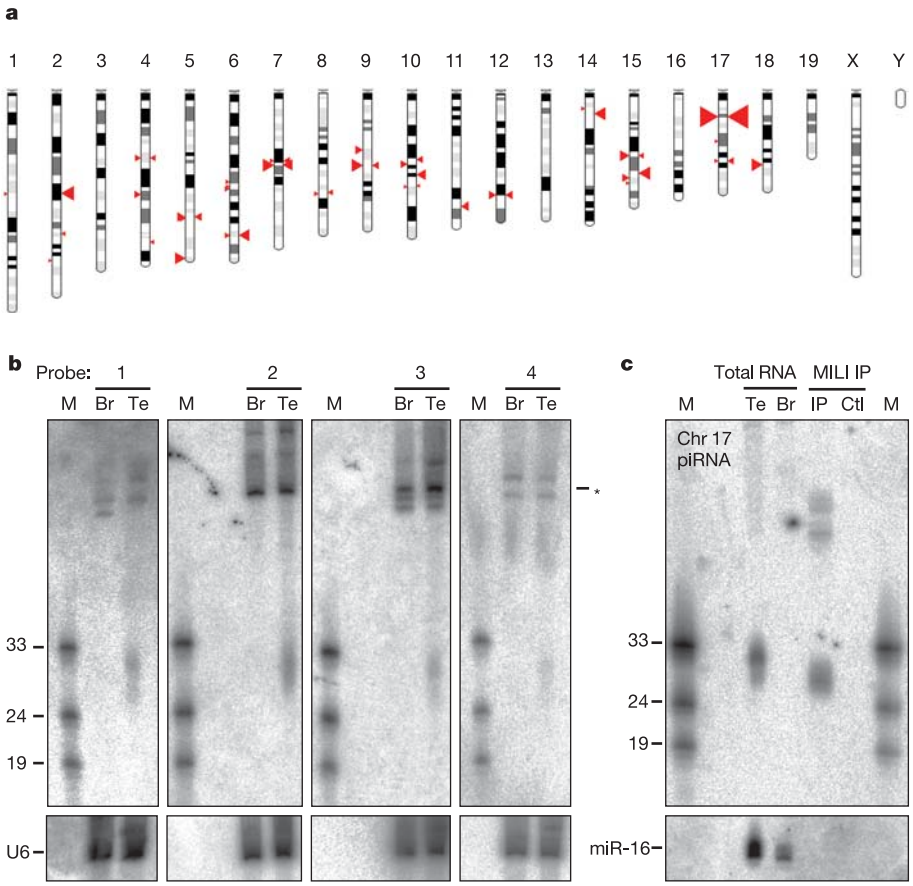


Figure 2 | MILI-interacting piRNAs are encoded in clustered genomic loci. a, Location of piRNA-encoding regions on mouse chromosomes. The size of the triangles is proportional to the number of cloned MILI-interacting sequences. The left or right position of the triangles indicates mapping of the clones to the minus or plus strand, respectively. **b**, **c**, Northern blot hybridization for mouse chromosome (Chr) 17 piRNAs using partially hydrolysed RNA probes against four distinct 500-nt regions (**b**) or a 27-nt oligodeoxynucleotide probe complementary to a single piRNA clone (**c**) (top panels). Mouse total RNA from adult testis (Te) or brain (Br) was examined, and the size and mobility of the RNA size marker (M) is specified. The asterisk marks tRNA cross-hybridization signals. Blots were re-probed with a U6 antisense probe (bottom panels). MILI-immunoprecipitated RNAs (IP) are absent in control (Ctl) immunoprecipitation experiments; miR-16 is detectable in total RNA from testis or brain, but is undetectable in the MILI-immunoprecipitated sample.

© 2006 Nature Publishing Group

this arrangement was caused by two closely spaced clusters: one that almost exclusively contained sequences mapped to the sense strand and the other containing sequences mapped to the antisense strand, suggesting bidirectional divergent transcription from a central promoter (Supplementary Table 5; Supplementary Fig. 2). Thus, although piRNAs resemble in size the rasiRNAs of fruitfly and zebrafish[10,12], they differ from rasiRNAs in several respects. First, piRNAs map to the genome in a highly strand-specific manner, in contrast to rasiRNAs that map to repeat regions in sense or antisense orientation as if randomly generated from long dsRNA precursors[10,12]. Second, piRNAs predominantly map to single genomic loci, whereas rasiRNAs map by definition to repetitive sites including transposable elements. In fact, the proportion of repeat elements is smaller within the piRNA regions than in the 100-kb flanking regions (29% versus 38%, P-value $< 2.2 \times 10^{-16}$).

The spacing between cloned piRNAs within each genomic region has no apparent pattern. To assess whether the piRNAs show any evidence of specific processing, we aligned the piRNAs whose loci are partially overlapping (52% of the clones in the piRNA regions) and evaluated the precision with which their 5′ and 3′ ends have been processed. We found frequent examples of piRNAs whose 5′ or 3′ ends coincide (Supplementary Tables 6 and 7), indicating that they are not random degradation products of long transcripts (Supplementary Fig. 3). Additionally, the 5′ end seems to be more precisely processed than the 3′ end, similar to what has been observed among miRNA clones[11]. To examine if piRNAs may be—like miRNAs—excised from dsRNA fold-back precursor structures, we investigated if any base-paired regions emerged in approximately 100 nt from each side of the piRNA. Although our computational approach was able to reveal the loop and stem regions of miRNA precursors, no clear hybridization pattern involving the piRNAs or the sequences flanking them was found (Supplementary Fig. 4). It is possible that long-range dsRNA structures or sequence-specific protein machinery are involved in guiding the maturation process. The strong preference for a 5′ uridine in piRNAs suggests the involvement of Drosha or Dicer RNase III, but additional structural or sequence-specific determinants are yet to be identified.

We further examined the processing of piRNA primary transcripts by northern blotting of total testis RNA using probes specific to four 500-nt regions within the largest piRNA cluster on mouse chromosome 17. All four probes antisense to cloned piRNAs detected a signal at 26–31 nt in testis, but not in brain, RNA (Fig. 2b). None of the sense probes yielded any signal, supporting the strand-specific accumulation of piRNAs (data not shown). Surprisingly, even a single oligodeoxynucleotide probe antisense to an individual piRNA was sufficient to detect a 26–28-nt-size signal in MILI-immunoprecipitated RNA and a 29–31-nt-size signal in total testis RNA (Fig. 2c). Oligodeoxynucleotide probes to sequences immediately flanking the isolated piRNA failed to produce a signal (data not shown), suggesting that the processing of piRNAs occurs in a directed fashion. Examination of the 5′ and 3′ ends of a 30-nt cloned piRNA by rapid amplification of cDNA ends (RACE) showed that the 5′ ends of the piRNA RACE clones were invariant in both libraries, whereas the 3′-end clones were truncated by 2 nt in the MILI immunoprecipitate small-RNA library (Supplementary Fig. 5). It is possible that the more abundant 29–31-nt fraction represents an intermediate processing product or that it corresponds to small RNAs interacting with another Piwi protein expressed in testis. Re-probing of the RNA blot for the ubiquitously expressed 22-nt miR-16 produced a miRNA signal for testis total RNA but none for MILI-immunoprecipitated RNA, indicating that MILI was specifically loaded with piRNAs but not miRNAs.

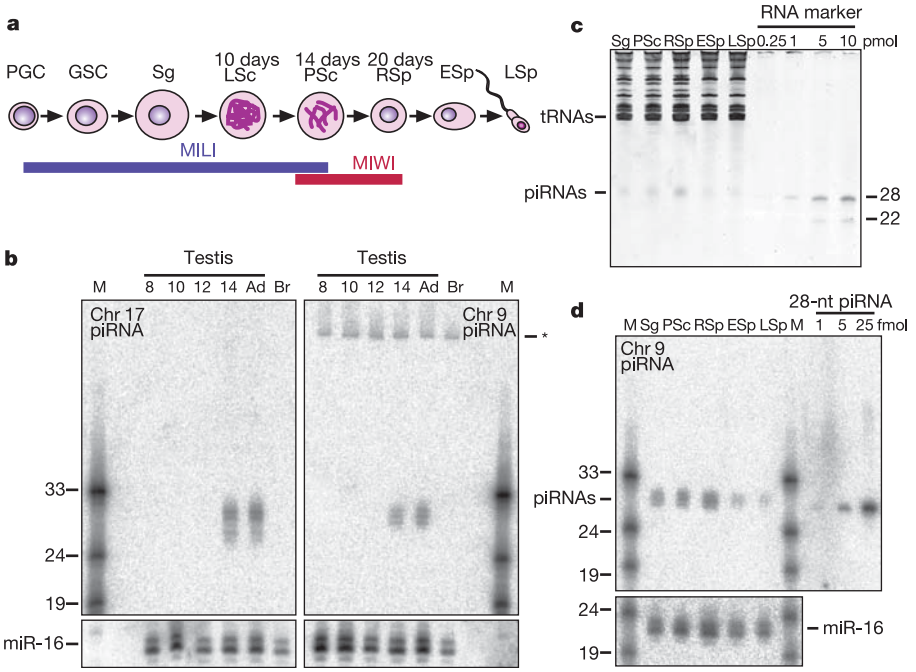


Figure 3 | Temporal expression of piRNAs during mouse spermatogenesis. **a**, Sketch of mouse spermatogenesis with the temporal expression patterns of MILI (blue) and MIWI (red) indicated. Abbreviations: PGC, primordial germline cells; GSC, germline stem cells; Sg, spermatogonia; LSc, leptotene spermatocytes; PSc, pachytene spermatocytes; RSp, round spermatids; ESp, elongating spermatids; LSp, late spermatids. **b**, Northern blot hybridization of testis total RNA from 8-, 10-, 12- and 14-day-old newborn and 3-month-old adult (Ad) mice and brain (Br) total RNA with oligodeoxynucleotide probes complementary to piRNAs (top panels) from chromosome (Chr) 17 (left panel) and chromosome 9 (right panel). The asterisk marks a cross-hybridization signal to a larger RNA also present in brain. Expression of miR-16 is monitored and serves as loading control (bottom panels). **c**, SYBR Green II staining of total RNA from elutriator-enriched male germ cells separated on a denaturing polyacrylamide gel. As reference RNA markers, two 22-nt and two 28-nt oligoribonucleotides of distinct sequences were loaded. **d**, Northern blot hybridization for samples shown in **c** using antisense probes specific for the piRNA cluster on chromosome 9 (top panel) and for miR-16 (bottom panel). A synthetic 28-nt chromosome 9 piRNA was loaded to quantify the amount of piRNA expressed in germ cells.

© 2006 Nature Publishing Group

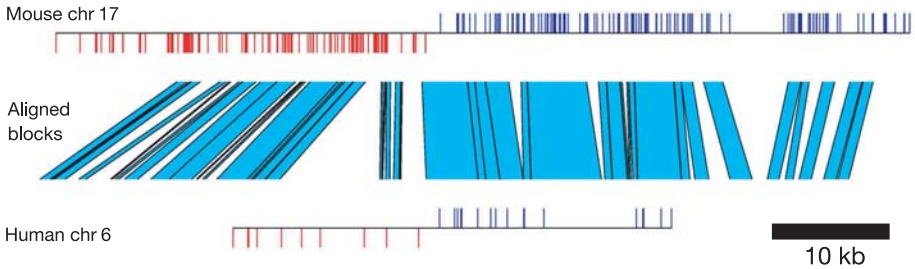


Figure 4 | Predominant mouse piRNA cluster and its orthologous cluster in human. Alignment view of the most highly expressed mouse piRNA cluster and its corresponding human orthologue. The positions of cloned sequences indicate divergent transcription from a central promoter in both species.

To obtain insight into the temporal expression of piRNAs during mouse spermatogenesis, we isolated total testis RNA at different time points of postnatal development. In mice, mitotically active spermatogonia represent the principal developing germ cells in testis up to day 6 after birth. Meiosis I is initiated on day 10, with germ cells reaching the preleptotene/leptotene, zygotene and early pachytene stages by days 10, 12 and 14, respectively[13] (Fig. 3a). MILI is expressed in male germ cells from primordial germ cells until the pachytene stage of meiosis[8], whereas MIWI is expressed in pachytene-stage spermatocytes and round spermatids[9]. Northern blotting of testis total RNA for two distinct piRNAs from mouse chromosome 9 and 17 revealed piRNA accumulation starting at day 14, when the first spermatocytes reach the pachytene stage (Fig. 3b). To assess the presence of piRNAs in specific germ-cell types, RNA was isolated from cells enriched for different stages of spermatogenesis after elutriation purification. Notably, the 30-nt piRNAs were so abundant that they were visible by SYBR Green II staining, and were estimated to be present at about 1 million piRNA molecules per mouse spermatocyte or round spermatid (Fig. 3c). Quantitative northern blotting for an individual piRNA from mouse chromosome 9 indicated about 8,000 copies per pachytene spermatocyte and about 2,000 copies per haploid round spermatid (Fig. 3d). The piRNA level was reduced by about tenfold in RNA isolated from later stages of germ-cell development.

Mouse piRNA sequences are well conserved and cluster within the closely related rat genome. However, the alignments of the mouse genome with eight other genomes[14] indicated that piRNA regions are poorly conserved between more distant species (Supplementary Fig. 6a), and that conserved elements are present with similar frequency within piRNA clusters and within introns of protein-coding genes (Supplementary Fig. 6b). For seven of the ten mouse piRNA clusters that seem to contain a bidirectional promoter, we found short regions of homology with the human genome. Moreover, we found that the frequency of ESTs that overlap with these human genomic regions orthologous to the mouse piRNA clusters was 9–21-times higher compared with the representation of testis ESTs among all the GenBank ESTs (Supplementary Table 5).

To provide experimental support for human piRNAs, we prepared and sequenced 18–26-nt and 24–33-nt small-RNA libraries from human testis total RNA (Supplementary Table 1; Fig. 1c). The small-RNA composition shows the expected enrichment for 5′ uridine, especially in the longer-size library where a 5′ uridine bias is not introduced by the presence of miRNAs. Using the same clustering criteria as for the mouse sequences, we were able to define 14 human piRNA clusters. They, together, contain 8.5% of all the cloned human sequences, and 24% of the human clones that were not derived from other functional RNAs (Supplementary Tables 8 and 9). The divergently transcribed piRNA cluster with the strongest expression in mouse is orthologous to the divergently transcribed cluster with the strongest expression in human (Fig. 4), and two additional human regions that are orthologous to divergently transcribed mouse piRNA clusters are experimentally supported (Supplementary Table 5).

Although Piwi proteins were shown to be important for stem- and germ-cell development in different animals[6–9], the underlying biochemical pathways are unknown. The identification of piRNAs provides an important molecular link regarding the function of Piwi proteins. Given the timing of the maturational arrest in *mili* and *miwi* knockout mice at the pachytene spermatocyte[8] and the spermatid steps[9], respectively, it is conceivable that piRNAs and germline-specific Piwi proteins regulate the timing of meiotic and postmeiotic events through transcriptional and translational repression. Argonaute proteins have been implicated in diverse processes such as genome rearrangement in *Tetrahymena*[15] or heterochromatic silencing and chromosome segregation in fission yeast[16], and we are only beginning to develop an understanding of the molecular mechanisms mediated by the diverse group of Argonaute ribonucleoprotein complexes.

METHODS

A detailed description of the methods used in this study is provided in Supplementary Information. Germ cells were obtained from the seminiferous tubules of 3-month-old C57BL/6J male mice using centrifugal elutriation as described previously[17]. Immunoprecipitation of MILI ribonucleoprotein complexes from mouse whole-testis whole-cell lysates was carried out using affinity-purified anti-MILI-pepN2 antibody raised against the peptide VRKDREEPRSSLPDPS (amino acids 107–122). Nucleic acids that co-immunoprecipitated with MILI were isolated, cloned and sequenced as described[18]. Small RNAs from total RNA of mouse or human testes were gel-purified, cloned and sequenced[18–20]. Cloned small RNAs were mapped to the mm6 and hg17 assemblies of the mouse and human genomes, respectively. The functional annotation was done as described previously[12,18,21], except that we updated our data set of non-coding RNAs and included predicted miRNA sequences[22–24]. For the repeat annotation, we used the repeat-masker results from the UCSC database. Clones were considered to be repeat-associated if they overlapped by at least 15 nucleotides with an annotated repeat element. The assessment of cross-species conservation of piRNAs and piRNA clusters was based on phastCons[14] conservation scores. The coverage of piRNA clusters by phastCons[14] conserved elements was also compared to the coverage of coding sequences and intronic regions of mouse RefSeq mRNAs[25] by conserved elements. Northern blots were performed as described previously[20], and the preparation and sequences of reference standards and probes are provided in Supplementary Information.

Received 18 February; accepted 17 May 2006.
Published online 4 June 2006.

1. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
2. Mello, C. C. & Conte, D. Jr. Revealing the world of RNA interference. *Nature* **431**, 338–342 (2004).
3. Matzke, M. A. & Birchler, J. A. RNAi-mediated pathways in the nucleus. *Nature Rev. Genet.* **6**, 24–35 (2005).
4. Zamore, P. D. & Haley, B. Ribo-gnome: the big world of small RNAs. *Science* **309**, 1519–1524 (2005).
5. Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
6. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727 (1998).
7. Reddien, P. W., Oviedo, N. J., Jennings, J. R., Jenkin, J. C. & Sanchez Alvarado, A. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science* **310**, 1327–1330 (2005).

© 2006 Nature Publishing Group

8. Kuramochi-Miyagawa, S. *et al. Mili*, a mammalian member of *piwi* family gene, is essential for spermatogenesis. *Development* **131,** 839–849 (2004).
9. Deng, W. & Lin, H. *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* **2,** 819–830 (2002).
10. Aravin, A. A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5,** 337–350 (2003).
11. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116,** 281–297 (2004).
12. Chen, P. Y. *et al.* The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes Dev.* **19,** 1288–1293 (2005).
13. Bellve, A. R. *et al.* Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. *J. Cell Biol.* **74,** 68–85 (1977).
14. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–1050 (2005).
15. Mochizuki, K. & Gorovsky, M. A. Small RNAs in genome rearrangement in *Tetrahymena*. *Curr. Opin. Genet. Dev.* **14,** 181–187 (2004).
16. Grewal, S. I. & Rice, J. C. Regulation of heterochromatin by histone methylation and small RNAs. *Curr. Opin. Cell Biol.* **16,** 230–238 (2004).
17. Jenab, S. & Morris, P. L. Testicular leukemia inhibitory factor (LIF) and LIF receptor mediate phosphorylation of signal transducers and activators of transcription (STAT)-3 and STAT-1 and induce c-*fos* transcription and activator protein-1 activation in rat Sertoli but not germ cells. *Endocrinology* **139,** 1883–1890 (1998).
18. Pfeffer, S. *et al.* Identification of microRNAs of the herpesvirus family. *Nature Methods* **2,** 269–276 (2005).
19. Pfeffer, S., Lagos-Quintana, M. & Tuschl, T. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) 26.4.1–26.4.18 (John Wiley and Sons, New York, 2003).
20. Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12,** 735–739 (2002).
21. Pfeffer, S. *et al.* Identification of virus-encoded microRNAs. *Science* **304,** 734–736 (2004).
22. Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B. & Bartel, D. P. Vertebrate microRNA genes. *Science* **299,** 1540 (2003).
23. Berezikov, E. *et al.* Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120,** 21–24 (2005).
24. Legendre, M., Lambert, A. & Gautheret, D. Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics* **21,** 841–845 (2005).
25. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33,** D501–D504 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Brennecke, M. Landthaler, Y. Pei, K. Fejes Toth and A. Sewer for discussion and help with the manuscript. We also thank L. Mitchell and K. Hwang for expert assistance in germ-cell purification, M. Poy and M. Stoffel for providing mouse tissues, and D. Weir and R. Choksi for assistance with sequencing. This work was supported by a FRAXA Research Foundation postdoctoral fellowship to A.A., an NIH grant to P.M., NIH grants to T.T., and an SNF grant to M.Z.

Author Contributions A.A., S.P. and M.L.-Q. prepared the mouse, and P.L. and N.I. the human, testes small RNA libraries. A.A. recognized the presence of piRNAs, performed the MILI IPs, prepared the MILI-interacting small RNA library, and performed, together with N.I., the northern blotting analysis. S.K.-M. and T.N. produced, characterized and purified the MILI antibody. T.T. developed the concept of cloning from Ago/Piwi IPs. P.M. isolated germline cells. M.C., J.J.R. and J.J. performed the large-scale sequencing. The bioinformatic analyses of piRNAs were designed and carried out by D.G. and M.Z. with input from A.A. and T.T. The database of RNAs with known function was compiled by M.Z., R.S., P.L. and S.P., the software used for small RNA annotation was developed by M.Z., R.S. and C.S., and the manuscript was written by A.A., M.Z. and T.T.

Author Information Sequences of the piRNAs determined in this paper are given in Supplementary Tables 4 and 9. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.Z. (Mihaela.Zavolan@unibas.ch) or T.T. (ttuschl@rockefeller.edu).

© 2006 Nature Publishing Group

LETTERS

Three-dimensional structure of the myosin V inhibited state by cryoelectron tomography

Jun Liu[1], Dianne W. Taylor[1], Elena B. Krementsova[2], Kathleen M. Trybus[2] & Kenneth A. Taylor[1]

Unconventional myosin V (myoV) is an actin-based molecular motor that has a key function in organelle and mRNA transport, as well as in membrane trafficking[1]. MyoV was the first member of the myosin superfamily shown to be processive, meaning that a single motor protein can 'walk' hand-over-hand along an actin filament for many steps before detaching[2–4]. Full-length myoV has a low actin-activated MgATPase activity at low [Ca^{2+}], whereas expressed constructs lacking the cargo-binding domain have a high activity regardless of [Ca^{2+}] (refs 5–7). Hydrodynamic data and electron micrographs indicate that the active state is extended, whereas the inactive state is compact[8–10]. Here we show the first three-dimensional structure of the myoV inactive state. Each myoV molecule consists of two heads that contain an amino-terminal motor domain followed by a lever arm that binds six calmodulins. The heads are followed by a coiled-coil dimerization domain (S2) and a carboxy-terminal globular cargo-binding domain. In the inactive structure, bending of myoV at the head–S2 junction places the cargo-binding domain near the motor domain's ATP-binding pocket, indicating that ATPase inhibition might occur through decreased rates of nucleotide exchange. The actin-binding interfaces are unobstructed, and the lever arm is oriented in a position typical of strong actin-binding states. This structure indicates that motor recycling after cargo delivery might occur through transport on actively treadmilling actin filaments rather than by diffusion.

Expressed full-length myoV formed two-dimensional (2D) arrays with a hexagonal unit cell ($a = 653$ Å) and a repeating motif that resembles a six-petalled coneflower (Fig. 1a, b). The 'pistil' of the flower contains the motor and cargo-binding domains, and the 'petals' contain the lever arms and the initial segment of the S2 domain. The arrays were extensive but the coherent domains were small, and the poorly sampled diffraction made spatial averaging unproductive. We therefore used correlation averaging and multivariate statistical analysis (MSA) to determine the molecular structure (Fig. 1b–h). This approach allows individual repeats to be aligned with each other independently and to be grouped with self-similar repeats when structural heterogeneity is present[11]. MSA of the flower and molecular repeats (Fig. 1b) reveals details not easily visible in the original micrographs. The major image features can be accounted for by the two motor domains, their lever arms and the S2 domain. The lever arms show a zigzag structure caused by the different orientations of the calmodulins[12]. Between the lever arms, the S2 domain is clearly visible and asymmetrically placed.

The assignment of the motor domains relative to the lever arms could be done in two ways. The favoured 'outward' arrangement (see below) has the motor domains extending roughly along the line of their lever arms (Fig. 1d); the 'inward' arrangement has the motor domains facing each other (Fig. 1e). With the exception of the initial segment of the S2 domain, the various parts of the rest of the tail are not easily separated.

To establish that the arrays consist of folded conformations with the heads bent backwards towards their S2 domain (as opposed to extended conformations, in which S2 could have come from a petal in the adjacent flower), we used MSA on the extracted molecules but included the neighbouring molecule to the outside (Fig. 1f–h). The three class averages revealed variable placement of the neighbouring molecules (Fig. 1f, g), and one of the classes (Fig. 1h) showed no neighbouring molecule. For this class, the S2 domain is as prominent as in the other classes, establishing that the arrays indeed consist of myoV molecules in the folded conformation.

We used electron tomography for three-dimensional (3D) image computation with higher signal-to-noise averages obtained by MSA of the 3D molecular images. In the raw tomogram (Supplementary movie 1) the arrays were readily identified, and flower and molecular

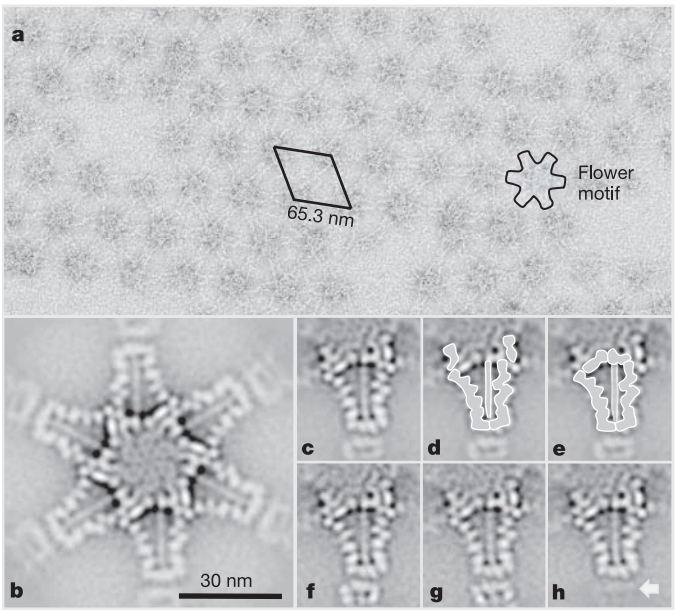


Figure 1 | Electron micrographs and averages of myoV. a, Negatively stained myoV 2D arrays. The 'coneflower' motif is outlined. **b**, Six-fold symmetrized average of 393 myoV coneflower repeats. **c**, Average of 2,358 myoV molecules. **d**, **e**, Two possible interpretations of the projection superimposed on the average: **d** is the outward arrangement and **e** shows the inward arrangement. **f**–**h**, Three different class averages of molecules using a classification mask that included parts of the neighbouring molecule at the bottom. At the bottom the adjacent molecules are different in **f** and **g**, whereas in **h** there is no adjacent molecule (arrow).

[1]The Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306-4380, USA. [2]Department of Molecular Physiology and Biophysics, University of Vermont, Burlington, Vermont 05405-0068, USA.

© 2006 Nature Publishing Group

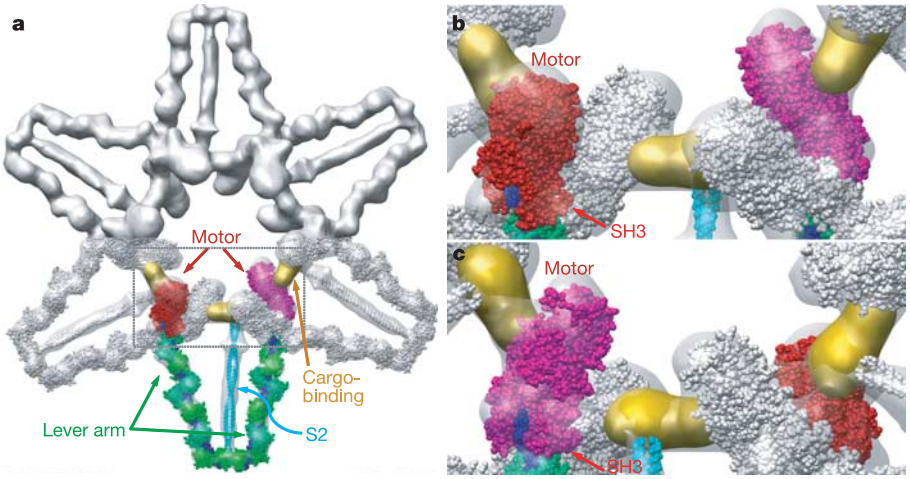


Figure 2 | Molecular arrangement within the 'flower' motif. a, At the top is an opaque surface rendering viewed from the solvent phase; at the bottom is a translucent surface view with the myoV atomic model rendered in space filling. The colour scheme for the bottom molecule is as follows: red and magenta, motor domains; green, light chains; blue, heavy chain component of the lever arm; cyan, S2 domain; yellow, cargo-binding domain density envelope; grey, adjacent molecules. **b**, **c**, Higher-magnification views showing the motor domain fit into the density envelope. **b**, View from the solvent phase; **c**, view from the monolayer side.

images could be extracted for alignment and processed as done for the 2D projections. A flower motif generated from the 3D class average with the most molecular images (Fig. 2a, and Supplementary movie 2) shows how six myoV molecules come together. One myoV head lies above (more into the solvent phase than) the head from the adjacent molecule, and their lever arms cross in the vicinity of the first calmodulin.

Array formation has the advantage of reducing conformational variability between individual myoV molecules but has the disadvantage that domain boundaries are sometimes ambiguous. To clarify the molecular boundaries we built an atomic model of the myoV motor domain and lever arm based on two X-ray structures[12,13] (Fig. 2). Two additional models for fitting were built with lever arms in the pre-powerstroke and 90° orientations based on smooth and scallop myosin S1 structures. For calmodulin, the 'closed' conformation[12] was a better fit than the 'open' conformation. Each of the initial models for the myoV heads was tested in both the 'inward' and 'outward' configurations using flexible fitting based on normal mode analysis[14]. The best fit (correlation coefficient of 0.82; Supplementary Methods and slideshow) was obtained for the structure with the post-powerstroke lever-arm orientation in the 'outward' configuration (Figs 1d and 2a).

The rod-like density can account for about 145 residues (about 21 nm) of the S2 domain (Fig. 2a), whereas the initial segment of S2 from sequence analysis consists of about 190 residues (about 28.5 nm)[15]. Some of the S2 sequence must be involved in reversing direction in the inhibited conformation, so the bulge at the end may be the PEST site. The remaining approximately 330 residues before the cargo domain are probably folded up and disordered within the pistil.

An extra density is present between the two motor domains (yellow in Fig. 2) that cannot be accounted for by the atomic model and is presumably the cargo-binding domain. It is positioned on the side opposite the motor domain's actin-binding surface (cyan in Fig. 3) and adjacent to loop 1 (green in Fig. 3). A decrease in the flexibility of loop 1 in myosin II (myoII) has been postulated to decrease ADP release rates[16]. Similarly, direct binding of the cargo-binding domain to loop 1 could stabilize its motion and decrease ADP release rates, thus accounting for the decreased ATPase activity of the folded monomer. The density has a volume corresponding to that predicted for a pair of cargo-binding domains and could be segmented as two separate densities of slightly different sizes. On the basis of the model building, these two closely juxtaposed densities come from different myoV molecules in the lattice. The 15-nm separation of the densities with respect to an individual myoV molecule is within the roughly 20 nm distance observed in electron micrographs of isolated myoV molecules[15].

Although the two myoV heads have different orientations within the arrays (Fig. 2), both have very similar, rigor-like conformations when docked on an actin filament (Fig. 4a, b). Each head can be docked without steric interference from the other head, the S2 domain or the cargo-binding domain. Binding of the inhibited conformation to actin was verified by decorating actin with myoV under conditions used to form the arrays (ADP + P_i). Electron micrographs of negatively stained specimens most commonly revealed extensive decoration (Supplementary Fig. S1). At lower myoV concentration, partial decoration could be achieved. Individual actin-bound molecules showed a triangular shape highly characteristic of our myoV structure (Fig. 4c). The main difference is a less angled lever arm, although this may be an artefact of drying. Our data

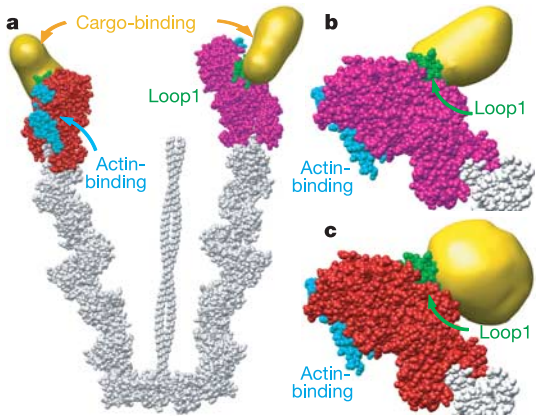


Figure 3 | Placement of the cargo-binding domain on the motor domain. Two cargo-binding domains are resolved that together have a peanut shape. The two motor domains are coloured red and magenta, and their actin-binding surfaces cyan. Loop 1 (residues 184–191) and the ATP-binding region (residues 424–430 and 228–235) are coloured green. The cargo-binding domain is positioned at loop 1. **a**, View from the same direction as Fig. 2a. **b**, Motor domain reoriented to show the relationship between loop 1, the cargo-binding domain and the actin-binding interface. **c**, Same view as **b** for the second motor domain.

© 2006 Nature Publishing Group

therefore indicate that the inhibited conformation might bind actin strongly.

Global conformational changes from extended active states to folded inactive states have now been observed in several molecular motors. Vertebrate smooth and non-muscle myoII[17], kinesin[18] and myoV[8–10] all form folded conformations with higher sedimentation coefficients and decreased ATPase activity. Although the overall strategy used by various motors is similar, the details of the inactive state are tuned to the particular motor, its molecular architecture and its functional role in the cell.

The mechanism of ATPase inhibition differs for myoII and myoV despite the folded form of their inactive structures. In myoV, ATPase inhibition occurs for both heads by the same mechanism, namely binding of the cargo domain to loop 1, a feature known to affect rates of nucleotide exchange[16]. In contrast, a key feature of myoII inhibition was the striking asymmetry of the interaction between the two heads, indicating that ATPase activity of each head might be inhibited by a different mechanism[19]. The myoV lever arm is in a post-powerstroke orientation; the molecule is therefore likely to bind strongly to actin (see Fig. 4). In contrast, the lever arm of myoII is in a pre-powerstroke conformation, namely M–ADP–P_i (ref. 19), indicating that it should bind weakly to actin, in accord with results from biochemical studies[17]. The switch that activates myoV is the binding of calcium and/or cargo[9]. Activation by calcium was observed directly from ATPase measurements, whereas activation by cargo is inferred from the structure because cargo binding should compete with and disrupt the intramolecular interaction between the cargo and motor domains. Phosphorylation of the regulatory light chain is the switch that extends the folded conformation of smooth muscle myoII[17].

The microtubule-based motor kinesin also undergoes a shift in sedimentation coefficient from about 6S to about 9S (refs 18, 20) that requires the cargo-binding domain. Artificial cargo can activate full-length kinesin, whereas shorter constructs are always active[21]. ATPase inhibition is due to slowed ADP release on the initial binding to microtubules, which keeps the microtubule-binding affinity weak[22].

The structures and properties of the inactive conformations of myoII and myoV indicate that there are likely to be unique mechanisms of motor recycling. MyoV is a barbed (+)-end-directed motor and once it delivers its cargo it has no obvious mechanism by which to return to its starting position to rebind cargo. Possible solutions would be simple diffusion or a set of recycling motors to carry the molecule back to its starting position. Both of these would require an inactive state with low affinity for actin, but our myoV structure indicates possible high actin-binding affinity.

An explanation may be found in the observation that, in brain, myoV co-localizes to regions such as growth cones[23] and dendritic spines[24], which have high rates of actin turnover indicative of active treadmilling[25]. We propose that actin treadmilling provides the recycling mechanism. MyoV cargo delivery would be akin to delivering a package by 'running up the down escalator'. Once the cargo has been delivered to the actin (+) end, myoV need only remain on the actin, which is facilitated by a conformation that inhibits ATP turnover but not actin binding. The treadmilling process on actin will eventually deliver the myoV to the actin filament (−) end, where new cargo binding can unfold and activate the molecule.

For this model to work, myoV must move at least twice as fast on actin as the actin filaments are growing. MyoV rates of movement on actin are of the order of 400 nm s^{-1} (refs 3, 15), whereas growth of the actin filament (+) end *in vivo* is about 10% of that by several measures[26]. Thus, myoV has the motor properties to 'run up the down escalator', and the binding properties to hang on for the return trip.

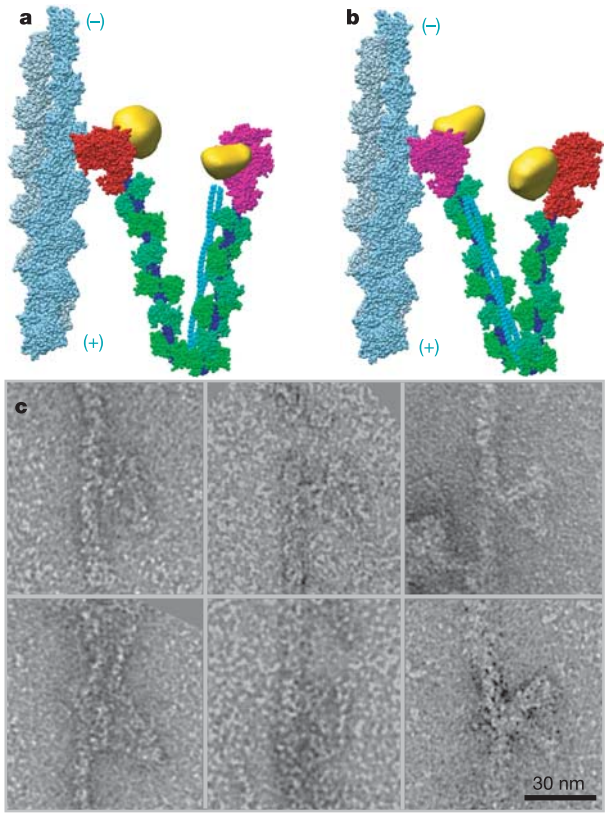


Figure 4 | Orientation of the myoV-inhibited structure on F-actin. F-actin strands rendered light blue and grey. Actin docking is based on ref. 30. **a**, Panel shows F-actin can bind only one myoV head at a time. The unbound head extends up towards the viewer. **b**, Same view direction as **a** but with the second head docked. **c**, High-magnification gallery of actin-bound myoV molecules. In some the S2 domain can be seen between the two heads. Drying in negative stain would flatten the structure onto a plane, which may account for some of the differences between the models and the micrographs.

METHODS

Expression, purification and 2D crystallization. A Flag-tagged, full-length murine myoV was coexpressed with a calmodulin mutant deficient in Ca^{2+} binding (CaMΔall) in *Spodoptera frugiperda* (Sf9) cells, and purified as described previously[9]. Crystallization, similarly to previous efforts[19], used a buffer consisting of 20 mM Na_2HPO_4, 80–100 mM NaCl, 2 mM $MgCl_2$, 1 mM ADP, 1 mM EGTA and 8–9% poly(ethylene glycol) 8000. 1,2-dilauroyl-*sn*-glycero-3-phosphocholine was from Avanti Polar Lipids. Other reagents were from Sigma-Aldrich. The concentration of MyoV was 0.2 μM.

Electron microscopy and image processing. MyoV 2D arrays were recovered from the lipid monolayer on 200-mesh copper grids covered with a reticulated carbon film. We used correlation averaging and MSA to obtain averaged images of the coneflower motif and its six individual myoV molecules. Coneflower repeats were selected interactively by using the Boxer module of the EMAN software package[27]. Extracted coneflower repeats were corrected for contrast transfer function (phase-flipped) by using the EMAN CTFIT module. Averaging was performed by first aligning the individual coneflower repeats, followed by the extraction and alignment of the six individual molecules. Several cycles of MSA and multireference alignment were used to produce class averages.

Electron tomography. Specimens were frozen as described previously[19], transferred into a Gatan 626 cryoholder and examined under low-dose conditions in a Philips CM300-FEG electron microscope operated at 300 kV. Eight single-axis tilt series were collected at ×43,200 magnification with a Tietz F224HD 2,048-pixel × 2,048-pixel charge-coupled device camera and the associated EM-MENU software (TVIPS). The pixel size at the specimen level was 5.56 Å. Tilt series of about 60 images were collected from −70° to +70° with the use of cosine-rule increments. Three different defoci (−5, −8 and −12 μm) were used for data collection. The cumulative electron dose of each tilt series was about 30 electrons $Å^{-2}$.

Tilt-series images were aligned by using marker-free alignment[28]. For defocus determination, small patches at constant *y* coordinate and therefore at the same defocus were extracted and their power spectra were summed incoherently.

© 2006 Nature Publishing Group

Defocus itself was determined with CTFIT[27]. Focus gradient correction[29] was applied to each image taken at more than 30° tilt, and a Wiener filter defocus correction was used for images with less than 30° tilt. The defocus-corrected images replaced the low-pass-filtered micrographs for further alignment refinement. Tomograms were computed by weighted back-projection.

3D average and correspondence analysis. Similarly to the 2D averaging, 3D averaging was performed in two steps. Coneflower repeats (200 pixels × 200 pixels × 80 pixels) were extracted from the cryotomograms, aligned in three dimensions and the six myoV molecules extracted (60 pixels × 80 pixels × 50 pixels). After all the myoV molecular volumes had been aligned to a single reference, they were classified with the use of hierarchical ascendant methods[11]. The 6,070 3D volumes were subjected to repeated cycles of multi-reference alignment and MSA until the signal-to-noise ratio no longer improved. In the final cycle, four class averages were generated; however, three of these, totalling 4,029 volumes, were judged to be very similar and were combined. The resolution of this average determined by Fourier Shell Correlation was 24 Å at the 0.5 cutoff. The 3D flower motif was generated by applying six-fold symmetry to the myoV molecule average.

Atomic model building. We constructed a myoV atomic model based on the high-resolution X-ray structure of the myoV–ADP motor domain plus the essential light chain structure (Protein Data Bank (PDB) accession number 1W7I)[13]. The rest of the lever arm was modelled on the X-ray structure of the calmodulin-like myosin light chain (Mlc1p) binding to the IQ2 and IQ3 motifs of yeast myosin Myo2p (PDB 1N2D)[12]. By repeating the IQ2 and IQ3 structures we built the IQ4, IQ5 and IQ6 complex. Because the 2D arrays were grown in ADP plus P_i, we tested the possibility that the lever arm was positioned in other orientations by rebuilding this structure at heavy-chain residue Ala 696 to conform to two other X-ray structures, PDB 1BR1 and 1QVI. These models were either poor fits to the density or produced structures that when docked on actin were flawed (Supplementary slideshow).

Received 15 January; accepted 15 March 2006.
Published online 16 April 2006.

1. Reck-Peterson, S. L., Provance, D. W. Jr, Mooseker, M. S. & Mercer, J. A. Class V myosins. *Biochim. Biophys. Acta* **1496,** 36–51 (2000).
2. Mehta, A. D. *et al.* Myosin-V is a processive actin-based motor. *Nature* **400,** 590–593 (1999).
3. Forkey, J. N., Quinlan, M. E., Shaw, M. A., Corrie, J. E. & Goldman, Y. E. Three-dimensional structural dynamics of myosin V by single-molecule fluorescence polarization. *Nature* **422,** 399–404 (2003).
4. Yildiz, A. *et al.* Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* **300,** 2061–2065 (2003).
5. Wang, F. *et al.* Effect of ADP and ionic strength on the kinetic and motile properties of recombinant mouse myosin V. *J. Biol. Chem.* **275,** 4329–4335 (2000).
6. Homma, K., Saito, J., Ikebe, R. & Ikebe, M. Ca^{2+}-dependent regulation of the motor activity of myosin V. *J. Biol. Chem.* **275,** 34766–34771 (2000).
7. Trybus, K. M., Krementsova, E. & Freyzon, Y. Kinetic characterization of a monomeric unconventional myosin V construct. *J. Biol. Chem.* **274,** 27448–27456 (1999).
8. Wang, F. *et al.* Regulated conformation of myosin V. *J. Biol. Chem.* **279,** 2333–2336 (2004).
9. Krementsov, D. N., Krementsova, E. B. & Trybus, K. M. Myosin V: Regulation by calcium, calmodulin, and the tail domain. *J. Cell Biol.* **164,** 877–886 (2004).
10. Li, X. D., Mabuchi, K., Ikebe, R. & Ikebe, M. Ca^{2+}-induced activation of ATPase activity of myosin Va is accompanied with a large conformational change. *Biochem. Biophys. Res. Commun.* **315,** 538–545 (2004).
11. Frank, J. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies* (Academic, San Diego, CA, 1996).
12. Terrak, M., Wu, G., Stafford, W. F., Lu, R. C. & Dominguez, R. Two distinct myosin light chain structures are induced by specific variations within the bound IQ motifs-functional implications. *EMBO J.* **22,** 362–371 (2003).
13. Coureux, P. D., Sweeney, H. L. & Houdusse, A. Three myosin V structures delineate essential features of chemo-mechanical transduction. *EMBO J.* **23,** 4527–4537 (2004).
14. Tama, F., Miyashita, O. & Brooks, C. L. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J. Struct. Biol.* **147,** 315–326 (2004).
15. Cheney, R. E. *et al.* Brain myosin-V is a two-headed unconventional myosin with motor activity. *Cell* **75,** 13–23 (1993).
16. Sweeney, H. L. *et al.* Kinetic tuning of myosin via a flexible loop adjacent to the nucleotide binding pocket. *J. Biol. Chem.* **273,** 6262–6270 (1998).
17. Sellers, J. R. Regulation of cytoplasmic and smooth muscle myosin. *Curr. Opin. Cell Biol.* **3,** 98–104 (1991).
18. Hackney, D. D., Levitt, J. D. & Suhan, J. Kinesin undergoes a 9 S to 6 S conformational transition. *J. Biol. Chem.* **267,** 8696–8701 (1992).
19. Wendt, T., Taylor, D., Trybus, K. M. & Taylor, K. Three-dimensional image reconstruction of dephosphorylated smooth muscle heavy meromyosin reveals asymmetry in the interaction between myosin heads and placement of subfragment 2. *Proc. Natl Acad. Sci. USA* **98,** 4361–4366 (2001).
20. Stock, M. F. *et al.* Formation of the compact confomer of kinesin requires a COOH-terminal heavy chain domain and inhibits microtubule-stimulated ATPase activity. *J. Biol. Chem.* **274,** 14617–14623 (1999).
21. Coy, D. L., Hancock, W. O., Wagenbach, M. & Howard, J. Kinesin's tail domain is an inhibitory regulator of the motor domain. *Nature Cell Biol.* **1,** 288–292 (1999).
22. Hackney, D. D. & Stock, M. F. Kinesin's IAK tail domain inhibits initial microtubule-stimulated ADP release. *Nature Cell Biol.* **2,** 257–260 (2000).
23. Wang, F. S., Wolenski, J. S., Cheney, R. E., Mooseker, M. S. & Jay, D. G. Function of myosin-V in filopodial extension of neuronal growth cones. *Science* **273,** 660–663 (1996).
24. Espreafico, E. M. *et al.* Primary structure and cellular localization of chicken brain myosin-V (p190), an unconventional myosin with calmodulin light chains. *J. Cell Biol.* **119,** 1541–1557 (1992).
25. Star, E. N., Kwiatkowski, D. J. & Murthy, V. N. Rapid turnover of actin in dendritic spines and its regulation by activity. *Nature Neurosci.* **5,** 239–246 (2002).
26. Forscher, P. & Smith, S. J. Actions of cytochalasins on the organization of actin filaments and microtubules in a neuronal growth cone. *J. Cell Biol.* **107,** 1505–1516 (1988).
27. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).
28. Winkler, H. & Taylor, K. A. Accurate marker-free alignment with simultaneous geometry determination and reconstruction of tilt series in electron tomography. *Ultramicroscopy* **106,** 240–254 (2006).
29. Winkler, H. & Taylor, K. A. Focus gradient correction applied to tilt series image data used in electron tomography. *J. Struct. Biol.* **143,** 24–32 (2003).
30. Rayment, I. *et al.* Structure of the actin–myosin complex and its implications for muscle contraction. *Science* **261,** 58–65 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Winkler for his guidance in focus gradient correction and 3D volume classification. We also thank the National Institutes of Health Research Resource for MMTSB and C. L. Brooks 3rd for making NMFF available. This research was supported by grants from the National Institutes of Health (to K.M.T. and K.A.T.).

Author Information The atomic coordinates of the final myoV atomic model have been deposited in the Protein Data Bank with accession number 2DFS. The density volume for the flower motif shown in Fig. 2 has been deposited in the European Bioinformatics Institute under accession code EMD-1201. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.A.T. (taylor@bio.fsu.edu) or K.M.T. (trybus@physiology.med.uvm.edu).

© 2006 Nature Publishing Group

LETTERS

The cargo-binding domain regulates structure and activity of myosin 5

Kavitha Thirumurugan[1], Takeshi Sakamoto[2], John A. Hammer III[3], James R. Sellers[2] & Peter J. Knight[1]

Myosin 5 is a two-headed motor protein that moves cargoes along actin filaments[1,2]. Its tail ends in paired globular tail domains (GTDs) thought to bind cargo[3]. At nanomolar calcium levels, actin-activated ATPase is low and the molecule is folded. Micromolar calcium concentrations activate ATPase and the molecule unfolds[3–6]. Here we describe the structure of folded myosin and the GTD's role in regulating activity. Electron microscopy shows that the two heads lie either side of the tail, contacting the GTDs at a lobe of the motor domain (~Pro 117–Pro 137) that contains conserved acidic side chains, suggesting ionic interactions between motor domain and GTD. Myosin 5 heavy meromyosin, a constitutively active fragment lacking the GTDs, is inhibited and folded by a dimeric GST–GTD fusion protein. Motility assays reveal that at nanomolar calcium levels heavy meromyosin moves robustly on actin filaments whereas few myosins bind or move. These results combine to show that with no cargo, the GTDs bind in an intramolecular manner to the motor domains, producing an inhibited and compact structure that binds weakly to actin and allows the molecule to recycle towards new cargoes.

Myosin 5 has two heavy chains, each contributing an amino-terminal motor domain, an α-helical lever stabilized by six sequentially bound calmodulin light chains, a coiled-coil region that dimerizes with the other heavy chain to form the tail and finally the GTD[3] (Fig. 1o). The unregulated high activity of myosin 5 heavy meromyosin (HMM) implicates the GTD in downregulation of activity at nanomolar calcium. Our initial images of the inhibited state showed a compact conformation, but identification of substructure was ambiguous[4]. The structures seen by other groups were less compact[5,6]. To understand the mechanism of regulation of cargo transport by myosins requires an understanding of the structure and properties of the inhibited state. Therefore we have studied the structure of folded myosin 5, and the GTD's mechanism of inhibition.

Single-particle image processing[7] of folded myosin 5 molecules negative-stained in ATP at nanomolar calcium levels reveals detailed substructure. Briefly, molecules abstracted from raw images (Supplementary Fig. 1) were aligned together, grouped into classes based on image features, and an average image was calculated for each class. Most features are consistent between the class averages (Fig. 1a–e), and identification of substructure is now straightforward (Fig. 1o). The two heads lie either side of an uninterrupted proximal 17 nm segment of the tail, with the head on the left of the tail more clearly delineated and closer to the tail. Within each head, calmodulin subunits of the lever domain can be seen, and the levers overlap at their junction with the tail. Overlap is more extensive in some classes (Fig. 1e), suggesting that levers are flexible. The tail length was highly variable in extended, rotary-shadowed molecules at high ionic strength[3]. In both raw and averaged negatively stained images of the folded state, we see only ~30% of the total predicted coiled coil[3]. Thus, portions of the coiled coil may be unstable and function as elastic linkers between motor and cargo[8]. Preliminary studies of molecules with no nucleotide or with ADP indicate that their structures are similar to those in ATP (Fig. 1h, i).

The motor domains have structural detail that can be reconciled with image averages of single heads and atomic structures. Visual inspection suffices for these comparisons because we have a single two-dimensional view of the molecule, and the features are distinct. The left head closely resembles a proportion of heads of myosin 5 HMM[9] and single heads (S1) (Fig. 1g) in ATP. It also matches well an atomic model built using the crystal structure of the myosin 5 motor domain in the 'post-rigor' conformation with an ATP analogue in the active site[10] (see Supplementary Methods; Fig. 1l). It does not match the pre-powerstroke conformation of myosin 5 HMM[9] or S1 (Fig. 1f) or a pre-powerstroke[11] atomic model oriented to show the same motor domain features (Fig. 1k), primarily because the motor is differently oriented on the lever. Differences in lever shape between folded molecules and models of heads are expected because folding may exert intramolecular strain and the lever is flexible[9] (see next paragraph and Supplementary Information).

The right head motor domain looks different. Its smaller size, strong stain exclusion and pronounced surrounding stain deposit indicate that it is approximately end-on. Its triangular shape is well matched by atomic models with the motor domain almost end-on (Fig. 1m, n). Although the appearance of the right lever resembles the pre-powerstroke model (Fig. 1n), the emergence point of the lever from the motor more accurately matches the post-rigor model (Fig. 1m). Moreover, right heads of molecules in ADP or nucleotide-free resemble those in ATP (Fig. 1h, i), yet are unlikely to occupy the pre-powerstroke state. This head is therefore probably in the same post-rigor conformation as the left head. Differences between images and the post-rigor model arise from lever flexibility[9] (see also Supplementary Information). Thus neither head appears to be in the pre-powerstroke conformation, despite this being the commonest conformation of heads of unregulated HMM in ATP[9].

The GTDs appear as two globules lying between the motor domains. A smaller lobe of the GTD, which touches the motor domain of the left head close to the lever, can be identified because it is absent from images of the head alone (Fig. 1a–e; compare with Fig. 1g; see also Fig. 1o). The length of each GTD (~7 nm) is similar to that of isolated, crystallized GTD from yeast Myo2 (8.8 nm; ref. 12). The consistent location of the smaller lobe indicates specific binding. Comparison with the atomic model (see Supplementary Information) shows that the GTD-binding site on the motor domain appears restricted to a specific segment, chiefly residues Pro 117–Pro 137 (pink in Fig. 1k–n and Fig. 2), and possibly extending to include His 138–Glu 154 (that is, helix E; ref. 13). The small lobe of

[1]Institute of Molecular and Cellular Biology, and Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK. [2]Laboratory of Molecular Physiology, NHLBI, National Institutes of Health, Bethesda, Maryland 20892-1762, USA. [3]Laboratory of Cell Biology, NHLBI, National Institutes of Health, Bethesda, Maryland 20892-1762, USA.

© 2006 Nature Publishing Group

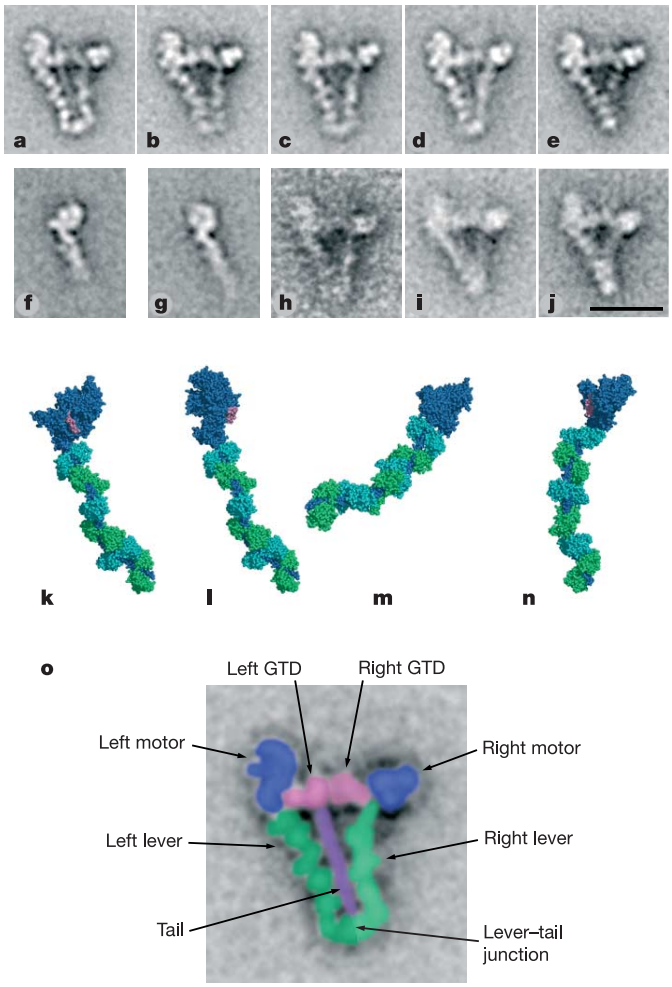


Figure 1 | Structure of switched-off myosin 5 and HMM-GTD complex. **a**–**e**, Averaged images of negative-stained, folded whole myosin molecules; 46–53 molecules per class. **f**, **g**, Averaged images of myosin 5 S1 stained in the presence of ATP; **f** shows the pre-powerstroke conformation (63 molecules), and **g** the post-rigor conformation (58 molecules). **h**, **i**, Averaged images of myosin in the presence of ADP or no nucleotide, respectively; 21 and 55 molecules. **j**, Averaged images of the complex of myosin 5 HMM and GST–GTD dimer; 28 molecules. The scale bar in **j** is 20 nm and applies to panels **a**–**j**. **k**–**n**, Atomic models of myosin 5 head; heavy chain shaded dark blue, the six calmodulins shaded alternately blue and green, and the putative GTD-binding region of the motor domain (Pro 117–Pro 137) shaded pink. **k**, Model using scallop motor domain containing ADP.vanadate[11] oriented to try to match left head of folded myosin in image averages. **l**, Model using myosin 5 motor domain[10] containing ADP.BeF_x oriented to match left head. **m**, Same model as **l**, oriented to match right head. Pro 117–Pro 137 lies behind the converter subdomain in this view. **n**, Same model as **k**, oriented to try to match right head motor domain appearance. Panels **k**–**n** were created using PyMOL (DeLano Scientific). **o**, Enlargement of **a**, coloured and labelled to show domains within folded myosin 5.

GTD associated with the right head is less prominent, but extends into the junction between motor domain and lever, which is the location of Pro 117–Pro 137 in right head models. Within the motor domain, Pro 117–Pro 137 abuts the entrance to the ATP-binding pocket, loop 1 and the β-bulge of the transducer[10], but is far from the actin-binding surface. GTD-induced conformational changes could readily modify nucleotide binding or hydrolysis and, through modifying the properties of the central seven-strand β-sheet, inhibit actin binding and stimulation of ATPase activity (see Supplementary Information). Thus each GTD binds at an equivalent site and acts allosterically, not by physically blocking binding to actin.

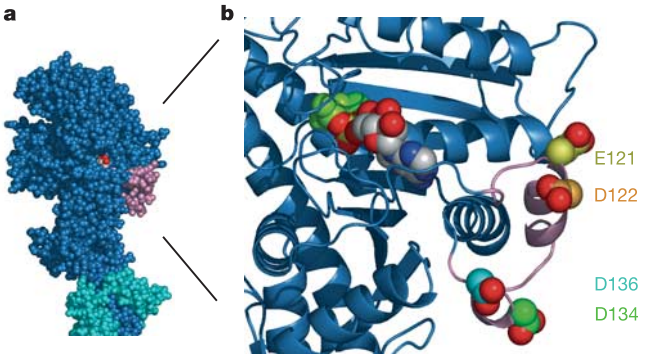


Figure 2 | GTD-binding region of myosin 5 motor domain. **a**, Motor domain region of the myosin 5.ADP.BeF_x complex. The heavy chain is shaded in dark blue, the proximal calmodulin in light blue and the putative GTD-binding region in pink. **b**, Enlargement of the GTD-binding region of the motor domain of **a** to show the spatial relationship to bound nucleotide. The polypeptide chain is shown in cartoon form, except for the four acidic residues (E121, D122, D134, D136) shown in spacefill with their carbon atoms and labels yellow-green, orange, green and light blue respectively and their carboxylate oxygens red. The bound MgADP.BeF_x is shown as a spacefilling model, with ADP shown with carbon silver-grey, oxygen red, nitrogen dark blue and phosphorus yellow, and BeF_x and Mg^{2+} green. The figure was created using PyMOL (DeLano Scientific).

We aligned amino acid sequences at the putative GTD-binding region of the motor domain for myosins 5a, b and c and compared this with other myosin classes (Supplementary Fig. 2). There are 3–4 acidic residues and no basic residues in any myosin 5a or 5b. The carboxy-terminal GDMDP sequence is conserved among vertebrate myosins 5, with conservative substitutions in invertebrates. Other myosin classes have non-conservative substitutions, indicating this may be a site of specificity for myosin 5 regulation. The acidic residues face outwards (Fig. 2b), suggesting ionic interactions with basic residues of the GTD. The densest cluster of basic residues in the GTD is five, within Arg 1800–Lys 1810. This is homologous sequence to helix 13 of the GTD of yeast Myo2, located at the distal end of the GTD[12], and therefore well-suited to match the morphology we see. Ionic interactions are consistent with the salt-sensitivity of folding[4–6].

The role of the GTD in regulation was studied using a glutathione S-transferase (GST)–47 kDa GTD fusion protein. GST is dimeric so the GTD is paired, as in myosin. $(GST–GTD)_2$ inhibits actin-activated MgATPase of myosin 5 HMM at nanomolar calcium levels with an inhibitory dissociation constant K_i of $0.53 \pm 0.14\,\mu M$ (s.d., $n = 5$), but not at micromolar calcium levels (Fig. 3a). GST_2 alone has no effect. Variable residual activity at saturating $(GST–GTD)_2$ partly derives from contaminant single-headed HMM[4]. This is supported by finding that $(GST–GTD)_2$ inhibits S1 more weakly ($K_i = 7.5\,\mu M$; Fig. 3b). The GTD is therefore sufficient to inhibit activity: the intervening sequence between the HMM coiled coil and the GTD is not required. S1 inhibition shows that paired heads are not required for inhibition, but do enhance the effectiveness. Electron microscopy shows that under inhibitory conditions, the HMM–$(GST–GTD)_2$ complex is a compact, triangular shape very similar to folded myosin 5 (Fig. 1j). Similarities include the difference between the shape of the two heads, and the shape and location of the GTDs, consistent with their binding the same sites as in myosin. $(GST)_2$, of mass (51 kDa) similar to a single GTD, is not visible, possibly being obscured by the central pool of stain.

It is a long-standing paradox that at nanomolar calcium levels, where ATPase is inhibited, myosin 5 bound to a coverslip moves actin better than at higher calcium concentration[3], and single-molecule motility assays show individual myosin 5 molecules moving along actin[14–16]. Discovery of the folded conformation at nanomolar calcium levels did not solve this puzzle, because folding echoes the

© 2006 Nature Publishing Group

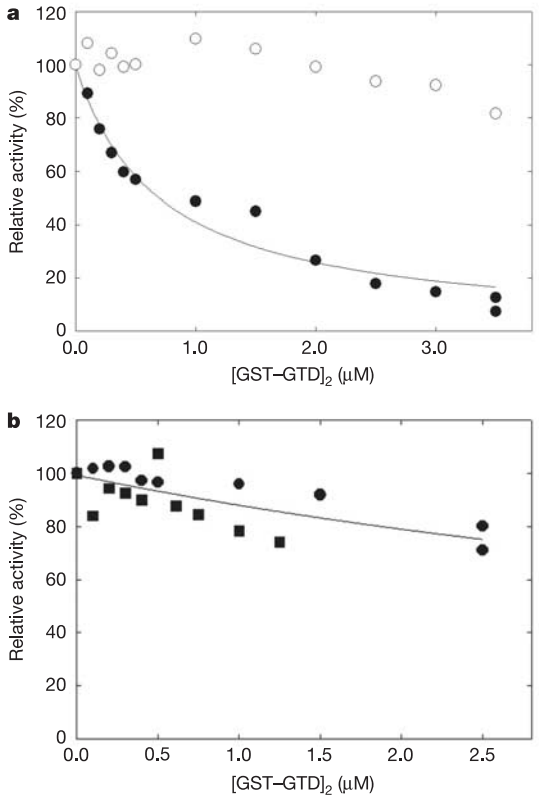


Figure 3 | Regulation of actin-activated myosin 5 ATPase activity by GST–GTD dimers. a, Myosin 5 HMM; solid circles at nanomolar calcium, open circles with 0.1 mM free calcium. The fitted line gives $K_i = 0.75\,\mu$M. **b**, Myosin 5 S1 at nanomolar calcium levels; squares and circles represent independent experiments. Activity is relative to ATPase in the absence of GST–GTD dimers. Ca^{2+} did not greatly affect the Michaelis constant (K_m) values for actin activation of HMM MgATPase which were generally in the range of 0.3–0.7 μM in the presence of 50 mM KCl.

'off'-state of myosin 2 and kinesin[17,18]. We addressed this paradox using two-colour TIRF (total internal reflection fluorescence) microscopy to measure simultaneously the single molecule motility of GFP–HMM and Cy3-myosin in mixtures interacting with the same actin filaments. Observing both species interacting with the same actin filaments overcomes difficulties of comparison between separate experiments. Without ATP, the filaments became heavily decorated with both proteins (Supplementary Fig. 3). With ATP, labelling was lighter with rapid movements of both proteins (see Supplementary Movies). The frequency of movements of each protein showed HMM movements (560 events per 100 pM HMM per min) were much more frequent than myosin movements (14 events per 100 pM myosin per min). This indicates that the great majority of myosin molecules do not move on actin, owing to being switched off. Including 2.5 μM $(GST–GTD)_2$ in the mixtures reduced HMM movements from 660 ± 50 to 317 ± 26 events ± s.d. per 100 pM per min, but myosin movements were unaffected (16 ± 6 to 15 ± 6 events per 100 pM per min), showing that the assay senses inhibition of activity. Myosins that move are probably either a small proportion of molecules that cannot fold or molecules that are unfolded when they collide with actin owing to the folded–unfolded equilibrium. Movement of a small proportion of myosin molecules explains the residual actin activation of myosin 5 ATPase[3,4,6] and the movements seen in earlier studies, and thus largely explains the paradox. *In vitro*, calcium disrupts the folded state and stimulates ATPase activity. Neither motor nor GTD binds calcium, so structural changes linked to calcium binding by calmodulins in the lever are the probable cause. In particular, dissociation of calcium-calmodulin from the lever, weakening it, is implicated in lowering motility[6,19]. This explains why the few active molecules at nanomolar calcium concentrations move actin better. Calcium-induced calmodulin dissociation may also disrupt the geometry required for inhibitory motor domain–GTD interactions, explaining how calcium stimulates ATPase activation by actin.

TIRF shows few, if any, long-lived (>1 s) stationary transient attachments of myosin to actin in ATP at nanomolar calcium concentrations. Moreover, electron microscopy of myosin–actin mixtures under these conditions shows detached, folded myosin and undecorated actin (Supplementary Fig. 4a). In contrast, with no nucleotide or with ADP or ADP + phosphate, TIRF shows heavy decoration, and electron microscopy shows that bundles of myosin-decorated actin filaments are immediately formed in solution (Supplementary Fig. 4b–d). Thus the folded molecules in ATP bind only weakly to actin with a fast 'off' rate, but under other conditions they bind more tightly.

The mechanism of myosin 5 regulation we have described differs from myosin 2, which uses head–head and head–proximal tail interactions[17,20,21] and therefore has regulated HMM[20]. Myosin 5 has more similarities with kinesin, in which a GTD binds the motor domain, slows ADP release and reduces microtubule affinity[18].

The GTD binds cargo receptor proteins as well as the motor domain. Activation of myosin 5 *in vivo* may therefore entail receptor proteins binding to the GTD, weakening the GTD–motor domain interactions we have described and allowing the heads to interact with actin[22]. Our data indicate that the folded molecule binds weakly to actin, so would be able to diffuse freely in cells. After myosin 5 dissociates from cargo at the cell periphery, folding would allow recycling to bind new cargo elsewhere in the cell.

METHODS

Proteins. The GST–GTD fusion protein comprised the C-terminal 414 residues of myosin 5a, encompassing the entire GTD, which was amplified by polymerase chain reaction (PCR) with Ultra High Fidelity Pfu (Stratagene) with *Bam*H1/*Eco*R1 ends, cloned into pGex-4T1 (Amersham) and expressed and purified by standard techniques. Other protein preparations are detailed in Supplementary Methods.

MgATPase assay. Actin-activated MgATPase assays were performed as previously described[23] using 20–40 nM HMM or S1, 10 μM actin, 50 mM KCl, 2 mM $MgCl_2$, 1 mM ATP, 0.1 mM EGTA, 10 mM MOPS (pH 7.0), 40 units per ml lactate dehydrogenase, 1 mM phosphoenolpyruvate, 200 units per ml pyruvate kinase, 200 μM NADH, 2 μM calmodulin at 25 °C. $CaCl_2$, if added, was 0.2 mM.

Single-molecule motility assay. Single-molecule TIRF assays were performed as described previously[23], except that a two-colour observation system (Dual view system, Optical Insights) was used to simultaneously view both GFP–HMM and Cy3-labelled myosin. 6–75 pM GFP–HMM and 60–4,000 pM Cy3-myosin were added into a flow chamber in the presence of ATP and recorded at a frame rate of $2\,s^{-1}$. The number of moving spots on all actin filaments was counted in a $22 \times 44\,\mu m^2$ area on both channels. Further detail is given in Supplementary Methods.

Electron microscopy and image processing. Typically, 40 nM myosin, 480 nM calmodulin, 100 mM KCl, 1.5 mM $MgCl_2$, 20 μM ATP, 0.1 mM EGTA, 2 mM potassium phosphate, 10 mM MOPS, pH 7.0 at 20 °C was applied to a carbon-coated electron microscope grid and immediately stained with uranyl acetate[24]. In some experiments ATP was replaced by hexokinase + glucose treated ADP (1.25 mM) or the same plus potassium phosphate (25 mM), or myosin was treated with apyrase. Actin, when present, was 0.75 μM, pre-treated with hexokinase + glucose or apyrase as needed. HMM–GTD complex comprised 40 nM HMM, 60 nM GST–GTD dimer, 55 μM ATP in the above buffer. Images were processed using SPIDER as described[24]. Further detail is given in Supplementary Methods.

Received 24 March; accepted 3 May 2006.

1. Vale, R. D. Myosin V motor proteins: marching stepwise towards a mechanism. *J. Cell Biol.* **163**, 445–450 (2003).
2. Sellers, J. R. & Veigel, C. Walking with myosin V. *Curr. Opin. Cell Biol.* **18**, 68–73 (2006).
3. Cheney, R. E. *et al.* Brain myosin-V is a 2-headed unconventional myosin with motor-activity. *Cell* **75**, 13–23 (1993).
4. Wang, F. *et al.* Regulated conformation of myosin V. *J. Biol. Chem.* **279**, 2333–2336 (2004).

© 2006 Nature Publishing Group

5. Li, X. D., Mabuchi, K., Ikebe, R. & Ikebe, M. Ca^{2+}-induced activation of ATPase activity of myosin Va is accompanied with a large conformational change. *Biochem. Biophys. Res. Commun.* **315,** 538–545 (2004).
6. Krementsov, D. N., Krementsova, E. B. & Trybus, K. M. Myosin V: regulation by calcium, calmodulin, and the tail domain. *J. Cell Biol.* **164,** 877–886 (2004).
7. Frank, J. *et al.* SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
8. Schilstra, M. J. & Martin, S. R. An elastically tethered viscous load imposes a regular gait on the motion of myosin-V. Simulation of the effect of transient force relaxation on a stochastic process. *J. R. Soc. Interf.* **3,** 153–165 (2006).
9. Burgess, S. *et al.* The prepower stroke conformation of myosin V. *J. Cell Biol.* **159,** 983–991 (2002).
10. Coureux, P. D., Sweeney, H. L. & Houdusse, A. Three myosin V structures delineate essential features of chemo-mechanical transduction. *EMBO J.* **23,** 4527–4537 (2004).
11. Houdusse, A., Szent-Györgyi, A. G. & Cohen, C. Three conformational states of scallop myosin S1. *Proc. Natl Acad. Sci. USA* **97,** 11238–11243 (2000).
12. Pashkova, N., Jin, Y., Ramaswamy, S. & Weisman, L. S. Structural basis for myosin V discrimination between distinct cargoes. *EMBO J.* **25,** 693–700 (2006).
13. Cope, M., Whisstock, J., Rayment, I. & Kendrick Jones, J. Conservation within the myosin motor domain: Implications for structure and function. *Structure* **4,** 969–987 (1996).
14. Sakamoto, T., Amitani, I., Yokota, E. & Ando, T. Direct observation of processive movement by individual myosin V molecules. *Biochem. Biophys. Res. Commun.* **272,** 586–590 (2000).
15. Yildiz, A. *et al.* Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization. *Science* **300,** 2061–2065 (2003).
16. Forkey, J. N., Quinlan, M. E., Shaw, M. A., Corrie, J. E. T. & Goldman, Y. E. Three-dimensional structural dynamics of myosin V by single-molecule fluorescence polarization. *Nature* **422,** 399–404 (2003).
17. Wendt, T., Taylor, D., Trybus, K. M. & Taylor, K. Three-dimensional image reconstruction of dephosphorylated smooth muscle heavy meromyosin reveals asymmetry in the interaction between myosin heads and placement of subfragment 2. *Proc. Natl Acad. Sci. USA* **98,** 4361–4366 (2001).
18. Hackney, D. D. & Stock, M. F. Kinesin's IAK tail domain inhibits initial microtubule-stimulated ADP release. *Nature Cell Biol.* **2,** 257–260 (2000).
19. Nguyen, H. & Higuchi, H. Motility of myosin V regulated by the dissociation of single calmodulin. *Nature Struct. Mol. Biol.* **12,** 127–132 (2005).
20. Sellers, J. R. *Myosins* (ed. Sheterline, P.) (Oxford Univ. Press, Oxford, 1999).
21. Ankrett, R. J., Rowe, A. J., Cross, R. A., Kendrick-Jones, J. & Bagshaw, C. R. A folded (10 S) conformer of myosin from a striated muscle and its implications for regulation of ATPase activity. *J. Mol. Biol.* **217,** 323–335 (1991).
22. Li, X. D., Ikebe, R. & Ikebe, M. Activation of myosin Va function by melanophilin, a specific docking partner of myosin Va. *J. Biol. Chem.* **280,** 17815–17822 (2005).
23. Sakamoto, T. *et al.* Neck length and processivity of myosin V. *J. Biol. Chem.* **278,** 29201–29207 (2003).
24. Burgess, S. A., Walker, M. L., Thirumurugan, K., Trinick, J. & Knight, P. J. Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules. *J. Struct. Biol.* **147,** 247–258 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. A. Burgess for expert advice and assistance with image processing, F. Zhang for technical assistance and The Wellcome Trust for support. T.S. was supported by a fellowship from the Japanese Society for the Promotion of Science.

Author contributions Electron microscopy and image processing were performed by K.T., TIRF microscopy by T.S., ATPase assays by J.R.S., and protein preparations by T.S. GST–GTD was created by J.A.H., molecular models by P.J.K., sequence alignments by J.R.S. and P.J.K., and the paper by all authors. J.R.S. and P.J.K. contributed equally to this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.J.K. (p.j.knight@leeds.ac.uk).

© 2006 Nature Publishing Group

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

What makes a good PhD student? This is an intriguing question, and one that was addressed recently in a short article in *Naturejobs* (see *Nature* **441,** 252; 2006). Georgia Chenevix-Trench, a research fellow at the Queensland Institute of Medical Research in Australia, wrote the piece and discussed ways in which PhD students can make the most of their studies. The article has since become one of the most popular ever to be published on our website. As well as enjoying a large readership, the piece has sparked an unprecedented amount of feedback, with a number of postdocs adding their voices to the debate on how to make science more enjoyable. Getting to know your colleagues as people, not just as scientists, is a good place to start, they write. They add that savouring your surroundings beyond the lab makes the experience that much more enjoyable.

These comments took me back to a turning point in my own graduate work. I was walking beside Lake Mendota at the height of a Wisconsin winter, wrapped up warm to ward off the freezing temperature. It was too cold to focus on anything except putting one foot in front of another, and keeping my nostrils — the only part of my body exposed to the elements — from icing up. I recall stopping to look around me and noticing that the sky was azure, the air crystalline. All thoughts of teaching duties, research problems and writing jobs disappeared in puffs of frozen air filtered through the burgundy wool scarf I had wrapped round my face. I was completely in the moment — and totally outside my responsibilities. It was then that I saw with absolute clarity how to resolve my research problems: the way best way to analyse my data and how to present them visually.

That moment was, for me, the epitome of what I believe the goal of all scientists should be: a synthesis of efficacy and enjoyment. Taking time to savour your surroundings — whether that means colleagues, friends, art, architecture or nature — can help nurture you, whatever stage you are at with your career. And it can make your work worth doing.

Paul Smaglik, *Naturejobs* editor

CONTACTS
Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
Peter Bless
Classified Sales Representatives
Tel: +1 800 989 7718
East USA/Canada: Andrew Bennie
NIH/Maryland/New York/ Pennsylvania: Shelley Cohen

San Francisco Office
Classified Sales Representative:
Michaela Bjorkman
West USA/West Corp. Canada
225 Bush Street, Suite 1453
San Francisco, CA 94104
Tel: +1 415 781 3803
Fax: +1 415 781 3805
e-mail: m.bjorkman@naturesf.com

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock
***Naturejobs* online production:**
Catherine Alexander

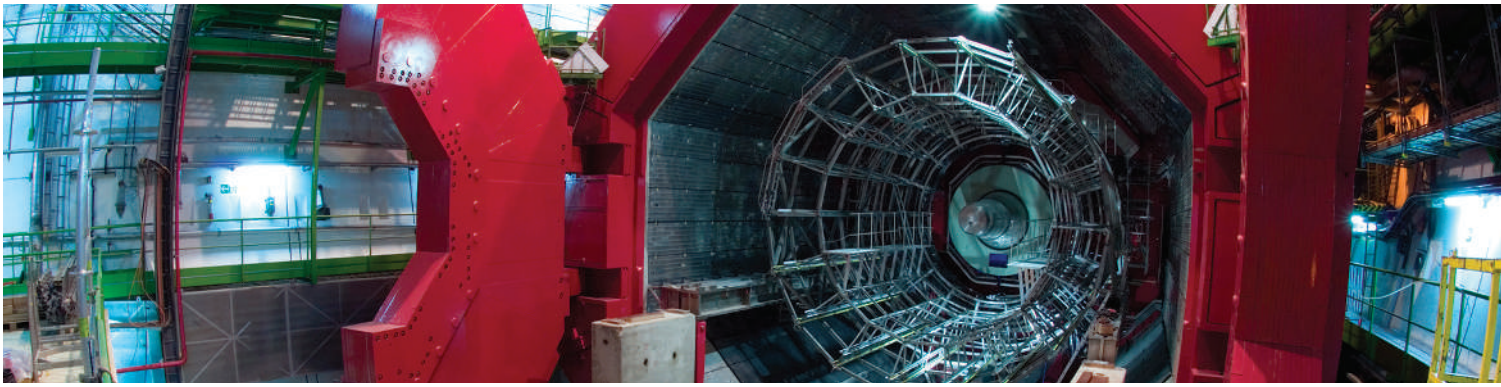
European Satellite Office
Patrick Phelan
e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

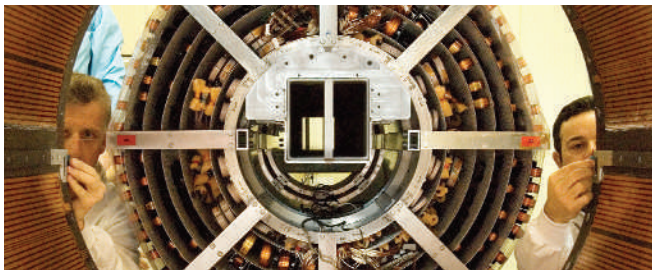
©2006 Nature Publishing Group

NATURE|Vol 442|13 July 2006



Physical exercise

The opening of the Large Hadron Collider in Europe will offer high-powered opportunities for particle physicists to decode the mysteries of the Universe. **Virginia Gewin** finds out more.



CERN

Physicists who dream of tracking down the building blocks of matter or of unravelling some of the secrets of string theory have their eyes on Switzerland. Construction of the Large Hadron Collider (LHC) is nearing completion at CERN, the European particle-physics lab near Geneva. Smashing subatomic particles together at unprecedented energies, the giant facility will offer physicists the chance to gain a deeper understanding of the early Universe.

Particle physics today needs more power — physically and analytically — than ever before. "New accelerators don't come online very frequently these days, and the LHC's size alone makes it important," says Per Osland, secretary for the European Physical Society's high-energy particle physics board.

The LHC is likely to reveal new physical phenomena, but in doing so it will generate reams of data, which will need to be mined. Four detector experiments will start up once the accelerator is switched on in November 2007. The two largest, ATLAS and CMS, are general-purpose detectors designed to measure the broadest range of signals, whereas the smaller detectors are designed to study, for example, heavy-ion collisions and the symmetry between matter and antimatter.

But instruments alone won't reveal the mysteries of the Universe. More than 7,500 international scientists are preparing the LHC's experiments — two of which will target one of the most alluring enigmas in high-energy physics: the Higgs boson. This tiny particle has never been measured but, if it exists, it would reveal the mechanism by which particles acquire mass. The working groups that have planned the experiments are now assigning specific tasks, such as detector installation and data analysis, to group members. No new groups are likely to be added, so joining an existing university group linked to one of the experiments is probably the best way to participate.

As relatively few open slots exist, having the right skills — especially in computation and data analysis — increases the odds of landing a position. The LHC data will be available via the Open Science Grid, an unprecedented network of distributed computing that will handle the 15 petabytes of data — the equivalent of about 15 million CDs — that will be generated every year. These data will be distributed among three tiers of computing centres, a federation of high-capacity global data grids. That alone will provide job opportunities at the interface of computing and physics.

Typically, physicists learn the computing skills they need to do their experiments. But with LHC data available on the grid, computer scientists who want to pick up physics could create data-crunching opportunities for themselves, says Miron Livny, facility coordinator of the Open Science Grid, who is based at the University of Wisconsin in Madison. "Success depends on how sophisticated you are in computing abilities, and that doesn't mean just learning a language," says Livny, adding that researchers must understand the principles that govern the computing infrastructure.

Per Osland is excited by the potential of the Large Hadron Collider (above) currently being finished near Geneva.

The right language

Experience using virtual neural networks to model data or multivariate analyses to mine data will also help, because the LHC physicists aren't sure exactly what they are looking for, says Michael Doser, a deputy head at CERN's physics department. Successful candidates need a broad range of experience, he adds. "We want high-flyers who have experience with computing, writing programs and analysis — not simply any one of these," he says. Some experience with detectors would be useful too, he says, as it would help researchers to spot whether a detector was working in the way simulations had suggested it would.

Some highly competitive career options are likely to be reinvigorated if the data reveal new phenomena. Theoretical physicists, for example, haven't been greatly in demand in particle physics in the past few years. But as the new data pour in, there may well be surprises that will open the way for theorists to guide future LHC experiments, says to Thomas Ludlam, senior scientist at Brookhaven National Laboratory in Upton, New York. Although CERN itself may provide few positions, thanks to financial constraints, some options exist.

©2006 Nature Publishing Group

Those wanting engineering and applied-science skills should pursue CERN's doctoral-student fellowships. Those wanting theoretical physics experience can look to the Marie Curie Early Stage Training in Theoretical Physics fellowships. Once the LHC is switched on, more of the roughly 50 two-year 'CERN fellow' postdocs filled each year will involve research.

Short-term positions exist at CERN for early-career scientists who have at least one postdoc under their belt. These are typically reserved for applicants from CERN's 20 member states, although exceptional US or Japanese candidates will also be considered. Non-member countries can fund their own positions at the same level, however. And undergraduates from around the world can take part in a three-month summer fellowship programme at CERN.

Balancing act

Not surprisingly, competition to work at the LHC is stiff, as top students swarm into the field. Although about 100 of CERN's scientists are due to retire during the next few years, most of these positions are likely to remain unfilled. CERN agreed not to recruit for the jobs in order to help repay the €300 million (US$380 million) loan from the European Investment Bank, which is being used to finish construction of the LHC.

Despite this, CERN will need to maintain expertise in electronics and engineering to get the accelerator and detectors up to full performance. Increasingly, however, the focus of accelerator physics is turning to the International Linear Collider (ILC), a proposed 30-kilometre collider that would smash electrons and positrons together.

Although the excitement surrounding the LHC is starting to eclipse the weaker accelerators that are already in use, those tried and tested facilities may still have much to offer (see 'The changing face of physics'). CERN is the prime destination for European particle physicists, but there are still opportunities in the United States at Fermilab in Batavia, Illinois, which is experiencing some difficulty maintaining staff levels. Fermilab, whose Tevatron is the most powerful accelerator now running, is luring international scholars with a fellows programme.

Existing accelerators — such as the Tevatron; the Stanford Linear Accelerator Center (SLAC) in Menlo Park, California; and the Relativistic Heavy Ion Collider at Brookhaven — are struggling to maintain relevance as they wind down. But their spokespeople say that they offer a mature environment that enables resourceful students to move on with their careers. "It would be silly to walk away from a working accelerator now to take a chance on the pitfalls a young machine may present," says Robert Roser, spokesman for Fermilab's Collider Detector Facility. Even though the Tevatron will stop operating soon after the LHC comes online, there will be years of data analysis to follow. Roser adds that students who want to fast-track their careers would be wise to pick a running experiment with a good track record.

The major collaborations required by the LHC mean that young participants could get lost in the crowd. "It can be difficult for an individual researcher or young physicist to have a big impact in such a large group," says John Womersley, director of particle physics at the Rutherford Appleton Laboratory near Oxford, UK.

And in the near term, graduate students and postdocs must balance their desire to make physics history against the potential for delay if they go to the LHC. Large accelerators sometimes require years to reach the stage of running smoothly, which could postpone the publications needed to move a young career forward. "Although no one doubts the LHC will work, the maiden voyage probably won't be straightforward," says Womersley.

The safe bet may be to finish up experiments at the older accelerators, where there will be no shortage of work. The challenge for Fermilab, for example, is to stretch personnel between Tevatron work, support for the LHC and the increased focus on the ILC, says Raymond Brock, a physicist at Michigan State University in East Lansing. Many US physicists plan to redirect much of their effort to the LHC once it comes online. The ATLAS experiment, for example, currently has the equivalent of 253 full-time employees. But Michael Tuts, the experiment's US programme manager based at Cornell University in Ithaca, New York, says he expects the number to reach 370 by 2009, largely as a result of redirecting effort from Fermilab and SLAC.

It might be tough going but, for many physicist-adventurers, the chance of jobs at older accelerators can't compete with the discovery potential of the LHC. "When was the last time anyone had the possibility of discovering a new dimension?" asks Doser. ■

Virginia Gewin is a freelance writer based in Portland, Oregon.

THE CHANGING FACE OF PHYSICS

Other facilities around the world engaged in high-energy physics now must face up to potential obsolescence thanks both to the power of the Large Hadron Collider (LHC) at CERN near Geneva, and plans being drawn up for the even more impressive International Linear Collider (ILC). Acknowledging that the LHC will outperform them, many facilities have sought to collaborate with CERN on data analysis while they expand into other areas (see *Nature* **435,** 713; 2005).

The Stanford Linear Accelerator Center in California, for example, is reinventing itself as a 'multipurpose' facility, in particular focusing on photon science and using its accelerator to generate X-rays that can be used for a number of experiments, including materials research and protein structure determination. DESY, Germany's main high-energy physics lab based in Hamburg, is similarly redirecting its efforts by building a 'free-electron' laser, which will generate X-rays for biological studies.

The Relativistic Heavy Ion Collider at Brookhaven National Laboratory in Upton, New York, is upgrading two of its largest detectors to continue its studies of what conditions were like in the early Universe. And Japan's high-energy physics lab, KEK, is building a proton accelerator for use in structural biology and materials research and is also pursuing its experiments on neutrinos. But it is Fermilab in Batavia, Illinois, that arguably faces the biggest challenge. Its particle collider, the Tevatron, is set to be shut down soon after the LHC comes on line. Pier Oddone, the lab's director, has already set his sights on making Fermilab the home of the proposed ILC (see *Nature* **435,** 728–729; 2005), which would undoubtedly boost the job outlook for particle physicists. **V.G.**

Thomas Ludlam believes that data from the Large Hadron Collider could open up fresh opportunities for theorists.

CERN

Web links

CERN Recruitment Service
https://ert.cern.ch/browse_www/wd_pds?p_web_site_id=1

Open Science Grid
lcg.web.cern.ch/LCG

Fermilab International Fellows programme
www.fnal.gov/pub/forphysicists/fellowships/international

International Linear Collider
www.linearcollider.org/cms

©2006 Nature Publishing Group

NATURE|Vol 442|13 July 2006

MOVERS

Jim Peacock, chief scientist, Canberra, Australia

2002–06: President, Australian Academy of Science, Canberra, Australia
1978–2003: Chief of Plant Industry, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, Australia

A delayed train helped Jim Peacock become one of Australia's leading life scientists. Arriving late at the University of Sydney, the education major found that his first choice of degree, the teaching of economics, was oversubscribed. He was advised to pursue his alternative options, botany and zoology, and is glad he did.

After gaining a PhD in genetics from the University of Sydney, he worked with noted *Drosophila* geneticist Ed Novitski at the University of Oregon in Eugene, then completed his training with an introduction to molecular biology at Oak Ridge National Laboratory in Tennessee. He returned to Australia to fulfil an obligation to his postdoc funder, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), but found that his work in plant science lacked the cutting-edge genetic tools being developed elsewhere for *Drosophila* research.

Peacock took the initiative and applied the molecular tools of *Drosophila* genetics to plant research. He became chief of the CSIRO's plant-industry division, and, with his colleagues, deciphered the sequence of corn (maize) transposons — genetic sequences that can be rearranged during replication.

Peacock held his CSIRO post for 26 years, and built the division up from 300 to 900 people. He also presided over the development of Australia's first — and so far only — transgenic crop. The genetically modified cotton is less susceptible to pests and is now used in several other countries. "It was a chance to relay fundamental discoveries into Australia's agriculture," he says.

After four years as president of the Australian Academy of Science, Peacock was chosen to become chief scientist of Australia. This position will allow him to guide government research investments, which he hopes will include more support for basic work. "I'm sympathetic to the pressures for results-based work, but there needs to be a proper balance," he says. He believes his experience with the controversies of genetic engineering have prepared him well to discuss hot topics such as exploring nuclear power as a source of energy.

Peacock doesn't duck controversial issues, notes CSIRO chief executive Geoff Garrett. "He has the interest, expertise and capability not only on plant science and agribusiness, but also across a broad range of science," he says.

Peacock's current passion is promoting science literacy in society. He's started taking advantage of his background in education to develop primary-school programmes and to interest young students in science. ■

Virginia Gewin

BRICKS & MORTAR

Science without the red tape

In 2000, in the wake of a period of political instability in Chile, a group of biologists and physicists came together to found an unusual institute. The Center for Scientific Studies (CECS) is based in the town of Valdivia, 800 km south of Santiago. One of the founders' initial hopes, says centre director Claudio Bunster, was that they might help redirect the efforts of Chilean military men and effect societal change by "forging links between civilians and the military through science". Now, for example, military officers fly converted anti-submarine aircraft for the centre's scientific missions to the Antarctic.

Bunster has an idealistic outlook, in part because his institute is small and relatively bureaucracy-free. When CECS biologists needed more time and money to finish a new genetics building, the institute's climatologists and glaciologists simply delayed their trip to the Antarctic for a year. This sort of compromise could be rather tricky in a typical university setting.

Major areas of focus at the CECS include biophysics, theoretical physics, and glaciology and climate change. Despite their disparate subject areas, CECS scientists often find ways to collaborate. One Antarctic mission run by glaciologists inspired biologists to search for extremophiles in a 4,000-metre-high salt lake in northern Chile.

The centre also boasts a milestone for South American biology: the continent's first certified transgenic mouse facility. Typically, such mice are imported, but CECS scientists have now bred their first transgenic mouse. Without proper sanitary and certified facilities, Chilean scientists can't design model animals or take a greater role in the community of mouse researchers, notes CECS biologist Marcelo Rubinstein. "We were completely lacking the technological capabilities," he says. The facility will house mice designed to model conditions ranging from obesity to neurological disorders.

The Chilean government provided one-third of the core funding for the CECS; the rest came from the World Bank's Millennium Science Initiative, the Fundación Andes and the Howard Hughes Medical Institute. The centre's yearly budget is just US$8 million.

In 1999, the CECS had 20 scientists, including graduate students; it now has 80, one-third of whom are foreigners. Bunster intends the centre to grow only another 20% at most. To retain the centre's soul, he plans to keep it small, agile and, as much as possible, bureaucracy-free. ■

Gene Russo

GRADUATE JOURNAL

Bowled over (but not out)

Next week are our assessments as first-year graduate students. I hope to make the transition from what Oxford calls 'probationer research student' to fully fledged DPhil status.

During the past few weeks I have written and rewritten a literature review, trying to condense many papers into a few thousand words. I have also struggled with computer graphics and animation while trying to make a slide show of my work that would neither send the audience to sleep nor make them wonder whether somebody had slipped something into their morning coffee.

During this time, I've had a welcome distraction in the form of the Promega Plant Sciences cricket team. I've had a great time playing as the 11th man (or, as I like to call it, 1st woman). Although some may find it patronizing, I enjoy being cheered for doing even the slightest thing right, such as passing the ball to the correct player. I wish I could enjoy the same feeling in the lab when I manage to get a clean blot or contamination-free agar plates.

Sometimes, though, lab life does mirror life on the pitch. You may not get the results you want, but you give it your all, you have fun, and the people with whom you share your experience make it all worthwhile. The trouble is, science is not a game: you need results. ■

Mhairi Dupré is a first-year PhD student in evolutionary developmental biology at the University of Oxford, UK.

©2006 Nature Publishing Group

NATURE|Vol 442|13 July 2006

The Republic of George's Island

One man against the elements.

Donna McMahon

On Tuesday afternoon, with the weather reports still forecasting hurricane-force winds, I hauled my decrepit fibre-glass dinghy down to Davis Bay and rowed out to try to talk that old throwback into coming ashore. I knew it was futile, but I felt like I had to make an attempt.

Westerly gusts drove whitecaps down the fetch of Georgia Strait. They rolled and crashed across the shallow shoreline, nearly overturning my clumsy boat. Icy spray slapped me as I strained at the oars, and within a minute I was soaked and achingly cold.

In the lee of the house, I tied up to a rusty trailer hitch. Decades ago, George had started driving deep steel pilings into the sandy soil around his house. He built a concrete retaining wall using old car and truck frames as rebar, and then piled up any other junk he could find for a breakwater. We kids watched with fascination while his neighbours, already besieged behind sandbag walls, shook their heads with a mixture of dismay and derision, but 30 years later, after three metres of sea level rise, his was the only original waterfront house remaining on the bay. And his sign, spray-painted on a full sheet of plywood nailed to the south side of the house, had become a local landmark:

"THE INDEPENDANT REPUBLIC OF GEORGE'S ISLAND — PISS OFF!"

I picked my way cautiously along the rough, algae-coated breakwater and positioned myself to one side of George's door before pounding on it and shouting.

"George, it's Logan. Let me in, OK?"

I couldn't hear much over the roar of wind and surf, but after several repetitions, a gruff voice bellowed back.

"Piss off!"

"I brought a bottle of rye."

There was a pause — probably George checking all his surveillance cams for evidence of an ambush — then the warped door opened a crack, revealing a wild tangle of grey beard and a bloodshot eye.

"You could've brought a 40-pounder," he grumbled. But he let me in.

So I sat for an hour at George's kitchen table in a stench of mildew and rotting carpet, trying to talk some sense into a guy with a rifle on his lap, chugging whisky straight from the bottle. Long ago he'd been a big man with a big gut, but he was well past 70 now, and his ancient, dirty clothes hung off him. I often wondered what he was eating these days. Any remaining cases of canned food must surely have rusted, and there wasn't much left in these plundered waters except barnacles and a few tenacious shore crabs. Even the glaucous gull population had plummeted.

The Pacific was slowly undermining George's foundations, and his cupboard doors hung askew, their melamine surface mildew-spotted and bulging with damp. Every room was crammed with piles and boxes of stuff — a mouldering graveyard of ancient flatscreen televisions, home appliances and power tools that had been sold or given away when people couldn't afford to run them any more. An ancient tungsten lightbulb, powered by a home-made solar system, illuminated the dirty kitchen and the illegal wood stove that George burned driftwood in.

JACEY

He waved the bottle at me, interrupting my warning about record-breaking winds and storm surges. I shook my head.

"No thanks. I don't drink."

"That just figures. Well, it's crap rye anyway."

Crap that I'd paid for out of my own pocket, I thought angrily, but I said: "Look, how about coming ashore until the storm is over?"

"So you can lock me up and then tear down the house like you did to Lawsons? Just how stupid do you think I am?" he muttered.

Pretty stupid, I thought. But mostly selfish. I'd met any number of old people who'd expected to be cocooned in consumer comfort all their lives, and who just couldn't get it through their heads that those days were over. They even had the gall to go crying to people like me who struggled to live off the scraps of their greed while busting our asses to rescue remnants of the ecosystem.

I wanted nothing to do with George, but my job as remediation team leader made it unavoidable. We were trying to re-establish intertidal zones on human-altered and polluted shorelines — filthy, frustrating work, the worst part of which was expropriating property that had fallen below sea level. The rolling easement laws were clear, but people nonetheless clung to their houses with blind tenacity. And George was the worst — sitting in his fortified pile and dumping raw sewage and garbage into my bay. Last summer, when I discovered he'd been digging for clams in the bivalve test site we'd spent three years building, I'd been ready to go out at low tide with a fire bomb.

"Last chance," I said finally. When George shook his head, I got up and walked to the door.

"Hey!" He rose unsteadily, and when I turned to meet his eyes, I saw fear. He knew.

Unexpectedly, he held out a grimy hand.

"Uh… Thanks, eh?"

I felt an unexpected lump in my throat as I shook his hand.

I didn't sleep much that night. I could hear the pounding surf even from my bedroom, a kilometre away from the beach. The next morning, the streets were littered with wind-shredded branches, broken glass, chimney bricks and roof tiles.

Down in the bay, waves foamed against a stump like a broken molar — chunks of concrete, twisted truck skeletons, and snapped off two-by-sixes. No house. No George.

I'd expected that. But I hadn't expected that I might miss him. ■

Donna McMahon lives in the small coastal town of Gibson's Landing (north of Vancouver, Canada), but spends much of her time in the twenty-second-century Pacific Northwest. Her degree is in history, which she stubbornly maintains is the field of the future as "there's more history every day".

©2006 Nature Publishing Group